Ritme and Rizumu

Studies in music cognition

Ritme and Rizumu

Studies in music cognition

een wetenschappelijke proeve
op het gebied van de Sociale Wetenschappen

Proefschrift

ter verkrijging van de graad van doctor
aan de Radboud Universiteit Nijmegen
op gezag van de Rector Magnificus prof. dr. C.W.P.M. Blom,
volgens besluit van het College van Decanen
in het openbaar te verdedigen op woensdag 15 november 2006
des ochtends om 10.30 uur precies

door

Makiko Sadakata
geboren op 29 mei 1976 te Osaka

Promotor

Prof. Dr. H. Bekkering

Co-promotor

Dr. ir. P.W.M. Desain

Manuscriptcomissie

Prof. Dr. W. Hulstijn
Prof. Dr. J.A. Michon, Leiden University, The Netherlands
Prof. Dr. K. Ohgushi, Kyoto City University of Arts, Japan

ISBN-10: 90-9021028-8
ISBN-13: 978-90-9021028-5

Contents

Chapter 1

Introduction

Human music performance is expressive; it is never rendered mechanical as the written score. Musical notes are stretched or delayed and loudness is increased or decreased. The fact that performance is not mechanical has been of a great interest to music cognition researchers. Music is a rich domain of study where diverse types of mental activity take place, involving both perceptual and motor processing at various levels. Therefore the search for the sources and the mechanisms of music behavior ultimately contributes to our understanding of general human cognition and behavior.

Many music cognition studies have looked into non-mechanical behavior in timing, loudness and pitch, especially in Western classical music. One of the reasons for the growth of research in this area is that these aspects of music performance are notated rather precisely in western classical music. In this line of work, expressiveness is often defined as deviation from the mechanical score information. Seashore (1938), one of the first to embark on music performance research, called this deviation 'artistic' . Later in the 80s, Bengtsson and his colleagues named it 'systematic variation' (Bengtsson & Gabrielsson, 1980; Gabrielsson, Bengtsson, & Gabrielsson, 1983), and today the word 'expressive' is common (as in 'expressive timing'). As such, the term is not used in the sense of 'what music expresses' but in the sense of the regularity and rules that underlie fluctuations in timing or loudness. The expressiveness of music performance in those contexts is quantitative; it can be measured, compared, and analyzed statistically. In this thesis, five independent studies will examine these quantitative music expressions, with reference to the loudness and particularly the timing dimension.

This thesis aims at increasing our understanding of music rhythm perception and music rhythm production. Several aspects are highlighted. First, we examine expression in rhythm production that derives from cultural experiences. Second, we consider the relation between rhythm in language and music, as language is a potentially influential factor on musical expression. Third, we propose a new way of relating rhythm perception and production to account for the differences between them noted in the literature. And finally, we address the effect of short-term training on rhythm perception and production, highlighting which aspects of their mental representation share a common basis. While very different approaches are taken in the individual chapters, the central concern throughout this thesis is to examine the effect of an individual's experience on musical behavior. The potential effect of exposure to one's native language as well as exposure to the occurrence of rhythmic patterns in the environmental music scene is examined in Chapters 2, 3, 4, and 5. The effect of short-term training is studied in Chapter 6.

This first chapter discusses music rhythm perception and production from three perspectives. First, we discuss the perception and production of expressive variations (further examined in Chapters 2, 4, 5, and 6) and review the relevant research findings. Second, we address the issue of cross-cultural effects on music expression (studied in Chapter 2, 3, and 5). And finally, a brief introduction to the learning of music performance studied in Chapter 6 is provided.

Perception and production of expressive variations

A range of studies have reported that trained individuals can perform a musical piece with consistent profiles (Gabrielsson, 1987; Palmer, 1989; Shaffer, 1984; Sundberg, Prame, & Iwarsson, 1996; Windsor & Clarke, 1997), and offer solid evidence that variations in music performance are not merely random fluctuations.

There are two types of expressive variation, one intentional and the other unintentional. Obvious evidence for intended expressive variation is that musicians can perform the same musical piece in different ways (Askenfelt, 1986; Kendall & Carterette, 1990; Nakamura, 1987; Palmer, 1989; Senju & Ohgushi, 1987; Windsor & Clarke, 1997). Even children who have not received formal music training can express different emotions using different articulations, pitch, dynamics and rhythms, for example (Adachi, 2004). Furthermore, Repp (Repp, 1992a, 1995b, 1997) showed that modulation range in musical expressions increases with musical experience. Nevertheless, this possible expressive space is not infinitely large, as expressions are at the same time constrained by the particular musical structure and performance style. If expression exceeds the 'boundary' of such a musical structure or style, the music sequence performed would be perceived in a different way. Thus, both freedom and constraint in musical expression are essential for artistic music performance (Timmers, 2002).

What are the main factors affecting expressive variation? One is certainly musical structure. There are many examples indicating that musicians shape their expression by making use of the elements offered by the musical structure. In other words, there is an intimate relationship between expression and structure, so that the main characteristics of the performance profiles of the same piece are very often consistent among different musicians. The same holds when a piece is performed by the same musician repeatedly over the course of several years (Shaffer, 1984). Additionally, musicians have trouble imitating an unintuitive profile (e.g., having to speed up at a position where one usually slows down, Clarke, 1993). Various models have been proposed in recent decades to predict expressive performance based on musical structure. For example, one of Todd's classical models (Todd, 1985) applies a parabola profile to predict timing and loudness according to the hierarchical structure of the musical phrases. This model appears to account very well for final phrase lengthening (Honing, 2003; Sundberg & Verrillo, 1980). The recent model proposed by Windsor, Desain, Penel and Borkent (2006) decomposes expression into meaningful units, such as timing profiles linked to metric levels and phrases. This model can predict small fluctuations during the phrases as well as a large final lengthening (for an overview of different models of rhythm and timing in music, see Clarke, 1999). The success reported in these approaches confirms that there is indeed an important correspondence between music structure and expression.

Another significant factor in musical expression is performance style. For instance, jazz musicians and classical musicians have been shown to perform very different renditions of the same piece (Ashley, 2002). Furthermore, individual musicians often develop their own performance style. These artistic styles are sometimes so characteristic that audiences can distinguish the performers from the sound alone. Although it is more difficult to deal with such artistic individual expressions scientifically, some studies have successfully outlined the individual characteristics of the performer as well as common characteristic aspects of music performance (Ashley, 2002; Collier & Collier, 2002; Gabrielsson, 1987; Repp, 1998, 1999a, 1999b). A broader overview of expressive performance is given in Gabrielsson (1999; 2003).

As has been mentioned above, the intended musical variations of an individual are to some extent systematic and rule-governed. This demonstrates that music performance is a highly cognitive activity, because the performer needs to have a well structured and systematic performance plan to realize it

(Sloboda, 1982). Furthermore, the fact that the listener can parse this expressiveness indicates that perception of expressive performance also involves complex cognitive processing.

Besides such intended expression, further fluctuation is always present in human music performance. For example, even highly trained musicians cannot perform music mechanically (Gabrielsson, 1974; Palmer, 1989; Penel & Drake, 1998, etc.). Several sources of such unintended expression are conceivable: variations introduced by motor noise, variations related to the music structure that cannot be freely chosen, and variations brought on by a perception and production bias (here the term 'production' is preferred over 'music performance' as we address more fundamental and general processes of producing acoustical signals),

Unlike motor noise that fluctuates randomly, unintended variations related to the music structure follows systematic patterns. For example, Gabrielsson (1974) reported recurrent unintended deviation patterns at regular positions in the rhythmic structure, such as timing profiles that systematically (short-long) deviate from the isochronous score timing. Several rhythm production (imitation) studies also report a tendency to simplify durational ratios of rhythmic patterns (Povel, 1981; Repp, Windsor, & Desain, 2002).

There is also a kind of perceptual bias associated with music structure which may contribute to obligatory expressive variations. A series of studies by Repp (Repp, 1992b, 1995a) showed that it is more difficult to notice the increment of inter-onset intervals at positions where musicians typically lengthen in real performance. Nakamura (1987) reported a similar phenomenon in loudness perception. Many listeners perceive a crescendo, an increase in the sound pressure level (SPL), at the end of the musical phrases while the SPL presented was constant or had even slightly decreased. This perceptual bias, which predicts the decline of the SPL at the end of the phrase, may contribute to the result; the crescendo was perceived because the presented SPL was louder than was expected by the listeners.

In the temporal domain, perceptual bias as well as some characteristic predispositions in production pose a complex issue in understanding the mechanism of music behavior, because the direction of the bias and the predisposition reported do not always coincide. In fact, the mutual linkages between perception and action are a widely investigated but still puzzling issue in human behavior. Recent cognitive theories, which do postulate mental representations, are manifold and usually focus on the link from perception to action or the other way around. A good overview of the different approaches is given in Hommel, Müsseler, Aschersleben, and Prinz (2001). Modern neurological work is now beginning to produce evidence for a biological underpinning of common representations in the form of mirror neurons that are activated both by the perception and the execution of an action (Rizzolatti, Fadiga, Gallese, & Fogassi, 1996, etc.). These discoveries may eventually shed light on topics such as imitative action. Though it is well known that imitation is primal and widespread, and that it is important for learning and development, how it is actually achieved is not yet understood.

In music, the link between perception and production can be quite automatic; for instance it can be hard not to move to a rhythm or to tap its beat. Furthermore, musical production, apart from being pleasurable, serves no other goal than to generate perceptions in the performer and in the audience, and in ensemble performance musicians have become experts at synchronizing their production with their perception of the production of others. Thus music, and rhythm and timing in particular, seems to be a natural and fruitful domain for the application of theories that link perception and production.

Acculturation effects on music perception and production

Just as musical instruments vary from region to region, so do rhythm, melody and scale. Several studies have demonstrated that people appreciate the musical structure from their own culture more than music that they are not familiar with (Drake & El Heni, 2003; Hannon & Trehub, 2005; Stobart & Cross, 2000). This suggests that an individual's music experience has a great impact on his or her current music perception and production. We call such an effect of exposure to environmental acoustical cues on music behavior the 'acculturation effect'.

Japanese culture has a unique history of acceptance of western music. Western tonal music was imported to Japan at the end of 19th century and made a significant impact. Nowadays western music has spread into every corner of Japan and is widely appreciated by both professional musicians and the general population. For example, Mito and Murao (1999) examined generational differences in acculturation to western music among the Japanese; it was easier for the younger generation to appreciate western tunes than the older one. Nevertheless, some authors have documented cases where Japanese musicians encounter problems of artistic expression when they perform western classical music (Saito, 1999; Shibata, 1983). It seems that even though they might have a technique good enough to perform difficult musical pieces, more authentic artistic expression – whatever it is – cannot be easily learned. The crucial question is why. Is there any systematic acculturation effect in music performance? What prevents some Japanese musicians from realizing authentic Western musical expression?

There has been little work on culturally specific expressive characteristics in the field of music cognition. One reason is that it is very difficult to measure and to generalize about cultural influence on music performance. Countless factors, such as history, education, social circumstances, and interactions among these factors, contribute to the creation of one's cultural environment. It is certainly a challenge to clarify which cultural aspects influence which features of musical expression.

Language is certainly one rather distinct feature shaping and shaped by culture. Growing attention to the relation between language and music has spawned many intriguing studies. One of the most influential works in this area is the Lerdahl and Jackendoff's Generative Theory of Tonal Music (1983). This theory highlights the formal similarity of language to music; the underlying structure of both domains can be descriptively expressed in a hierarchical manner, typically represented in a tree diagram. Other empirical cognitive studies have shown interesting parallels between the processing of speech and music information at various levels. For example, elements of both language and music are subjected to categorical perception (Cutting & Rosner, 1974; Desain & Honing, 2003; Eimas, Siqualand, Jusczyk, & Vigorito, 1971; Siegel & Siegel, 1977). An excellent overview of further parallels in the processing of language and music is given in MucMullen and Saffran (2004).

In addition to these parallels, there is also evidence of cross-domain influence; for example, an effect of statistical regularity in speech on musical behavior has been reported. Iversen, Patel and Ohgushi (2004) showed that their participants from different cultures tend to make different perceptual groupings of musical tones. Furthermore, these different patterns resembled the most frequent accent patterns in their mother tongues. Stevens (2004) also reported that people who speak a tonal language, in which lexical meaning is derived from the change of pitch in speech, are more sensitive to a task of musical pitch discrimination.

The critical question is whether cognitive paths are shared in language and music information processing. Research results have not yet yielded a clear answer. There has been evidence suggesting a commonality between domains in several empirical studies. For example, Serafine and her colleagues have shown that the memory traces of lyrics and melodies in songs are somehow integrated into one memory trace (Serafine, Crowder, & Repp, 1984; Serafine, Davidson, Crowder, & Repp, 1986). Other studies have shown that the rhythmic characteristics of a composer's mother tongue seem to be reflected in their music compositions (Patel & Daniele, 2003a; Wenk, 1987). Thus, the composer may develop a 'sense of rhythm' which is somehow shared by language and music. Moreover, the event related potential (ERP) components measured in electro encephalography (EEG) that are sensitive to syntactic violation of speech and music information processing have been shown to be similar (Osterhout & Holcomb, 1992, 1993; Patel, 1998). Additionally, Maess, Kelsche, Gunter, and Friederici (2001) have offered evidence that musical syntax is processed in Broca's area and its right-hemisphere homologue, which have long been believed to be responsible exclusively for the processing of language information. These neuroimaging data may suggest that these areas process any complex-rule based information like language and music.

On the other hand, there are many neuropsychological as well as neuroimaging studies indicating dissociation between domains. For example, although it is common that impairment in language ability (aphasia) and musical ability (amusia) occurs at the same time, exceptions do exist: there is aphasia without amusia, and amusia without aphasia (see Peretz, Gagnon, Hebert, & Macoir, 2004). These cases suggest that the brain areas responsible for information processing for speech and music do not completely overlap. In addition, Besson, Faïta, Peretz, Bonnel and Requin (1998) have shown that when both lyrics and tunes are violated in the song stimuli, the additive contribution of two ERP components which signal violated lyrics and music context was elicited. This suggests the independent processing of language and music information. Furthermore, Tervaniemi, Medvedev, Alho, Pakhomov, Roudas, and van Zuijen (2000) showed that the brain areas which respond to changes in phonemes and musical chords are different.

Clearly, continuing research is needed to improve understanding of the mixed results regarding association of cognitive paths in language and music information processing. Establishing theoretical frameworks, such as those provided by Peretz and Coltheart (2003) or Patel (2003), is important to the furtherance of research in this area.

Learning and Development

Acculturation effect is a slow learning process that takes place over many years. However, acquiring the skills to perform music is a more active process that involves practice and training. In fact, the importance of practice for the learning and development of music performance has been pointed out by several authors. Ericsson, Kranpe, and Tesch-Römer (1993) claim that one needs at least 10 years of deliberate practice to become an expert in most fields, including music performance. Sloboda, Davidson, Howe and Moor (1996) further investigated the association between practice and achievement in music performance, and concluded that formal effortful practice is a principal determinant of musical achievement.

These studies collected data from interviews that retrospectively evaluate the development of musical experience. This method is deemed powerful and, more importantly, claimed to be easier to utilize than recordings of students' development over time, and is commonly used in music education research studies (e.g., MacDonald & Wilson, 2006; Pitts, 2002; Woody, 2000). However, more experimental approaches can highlight learning and developmental aspects in music behavior with more reliability.

The studies in this line typically utilize a between-subject design and show the correlation between musical abilities and age or amount of musical training. Drake and her colleagues have looked into the effect of musical training on musical behaviors from various angles. These studies revealed, for example, that trained individuals can process more complex rhythms (Drake, 1993a, 1993c), tap more precisely (Drake, Penel, & Bigand, 2000a), and are aware of longer units, and therefore of higher hierarchical levels of structure in music (Drake, Penel, & Bigand, 2000b). As for the process of skill acquisition, Drake and her colleagues have emphasized the importance of development of cognitive aspects, including the ability of self monitoring, planning, and being aware of the musical structures (Drake & Palmer, 2000; Palmer & Drake, 1997).

While the approach that compares musicians and non-musicians is able to show which aspects of musical ability develop over the course of years of musical training, yet another experimental approach, the transfer of learning paradigm, can provide a more immediate picture of the process of learning and development. Transfer of learning refers to the situation, where, for example, one can apply knowledge previously acquired to a novel setting. This learning can be both implicit and explicit, where implicit learning means that participants are not aware of what they have learned whereas in explicit learning they are (but see Frensch & Rünger, 2003). The method can highlight a rapid type of learning process within individuals that takes place during every day practice.

While there have been a number of studies evaluating transfer of learning in various fields (Cohen, Ivry, & Keele, 1990; Heuer & Schmidt, 1988; Kilborn & Ito, 1989; Robertson, 2000; Scruggs & Mastropieri, 1988), there are few studies assessing this effect in the music cognition field (Meyer & Palmer, 2003; Palmer & Meyer, 2000). The approach can be applied to highlight various aspects. For example, one can examine the learning process in relation to one's experience, or the optimal relationship between amount of practice and improvement. The most important feature is that it can highlight which condition involved in the task makes use of the same representation or process, as transfer of learning is supposedly facilitated in such cases. Thus, this paradigm is certainly a powerful tool with which to gain further insights into the short-term effects of learning performance and perception of musical expression.

Preview

Chapter 2 ***A cross-cultural comparison study of the production of simple rhythmic patterns***

This chapter aims at increasing our understanding of the production of musical rhythms by examining the effect of culture. A cultural difference in rhythm production has been reported in a rich musical context (Ohgushi, 2002). The study investigates whether such cultural differences also exist in the production of simple rhythms. Dutch and Japanese percussion players participated in the experiment.

Chapter 3 ***Comparing rhythmic structure in popular music and speech***

This chapter addresses the question of why some musical characteristics appear to be culture-specific. We examine the relation between rhythmic characteristics in language and music and how this may explain the observed effect of culture. Patel and Daniele (2003a) showed a correlation between speech prosody and composed musical rhythm using an index that measures the contrastiveness of successive durations (nPVI). This correlation between music and speech is further investigated using a comparison between the rhythmic characteristics of English and Japanese popular music and English and Japanese speech. To evaluate the nPVI measure further, two factors which shape rhythmic structure were distinguished: 1) the distribution of the durations included in pieces of music, and 2) the sequencing of successive note durations. The question was which of these factors contributes more to the formation of the characteristics of rhythmic structures specific to the two cultures.

Chapter 4 ***The Bayesian way to relate rhythm perception and production***

This chapter proposes a method of relating the perception and production of musical rhythms. A Bayesian approach provides a new way of understanding the differences that have been shown to exist in the perception and production studies. It formalizes the perceptual competition among the different rhythmic categories in mental representations while assuming possible non-uniform a priori probabilities of the rhythmic categories; a certain category may have a better chance of being perceived than another. The method is tested on rhythm perception and production data taken from different studies in the literature. Independent prior probabilities of the rhythmic categories are derived from counts of patterns in corpora of musical scores, or from a theoretical measure of rhythmic complexity.

Chapter 5 ***Diversity and commonality in music rhythm perception and production: the influence of rhythmic complexity and culture***

This chapter elaborates the issues addressed in Chapters 2, 3 and 4. An acculturation effect on rhythm perception, and production is examined. The individual's familiarity with different rhythms is evaluated as well. One of the main aims of this study is to investigate the relation between the complexity of rhythms and diversity/commonality in their processing. Two complexity measures of musical rhythm are identified: a syncopation measure, which represents hierarchical complexity, and the nPVI, which represents serial complexity. The question is which of these two measures accounts better for the distributions of the data observed. The other aim is to apply the Bayesian approach proposed in Chapter 4 to relate rhythm perception and production, using a familiarity rating as a prior.

Chapter 6 ***Learning to perform simple musical rhythms with expressive variations***

The chapter addresses the issue of learning to produce different renditions of musical rhythms. We assess how quickly expressive variations in rhythms can be learned. The aim is to provide insights into the issue of which aspects of expressive music performance share a common basis. Certain rhythms may be more closely associated than others in terms of their mental representations, and therefore transfer of learning may be facilitated. This could provide evidence for categorization of a similar surface structure of the rhythm into a shared higher level mental representation related to the score.

Chapter 2

A cross-cultural comparison study of the production of simple rhythmic patterns

Abstract

It has been argued that Japanese and Western musicians give different impressions to the listener when performing western music. However, these claims are mostly based on subjective impressions and very few studies provide corroborative empirical evidence. The aim of the present chapter is to compare Japanese and Western musicians with regard to timing. Japanese and Dutch percussionists performed nine kinds of rhythmic patterns consisting of two intervals, under two conditions, in three tempi. There seemed to be a tendency for the Japanese participants to perform 4:1, 5:1, 1:4 and 1:5 patterns with a smaller duration ratio than instructed, although performance of patterns with ratios closer to 1 were similar between the participant groups.

Common music notation provides musical information such as the pitch and duration of intervals. In musical performance, musicians read the information from the musical score and convert it into sound sequences. Measurements of musical performances have been conducted since the beginning of the 20th century, e.g. by Seashore and his colleagues (Gabrielsson, 1999; Seashore, 1938), revealing systematic deviations from the music notation in the performed sequences. In later work, the differences between performed time intervals and their idealized counterparts in the score were shown to be linked to, and convey, the musical structure (Clarke, 1988; Sloboda, 1985). One may wonder how deeply rooted these processes are in human cognition and whether the relations found are valid across cultural boundaries.

There are numerous possible factors that characterize culture, such as language, social structure and history. A factor that is likely to have an influence on perception and production of rhythm is language: Japanese, for instance, is hypothesized to be a mora-timed language in which all 'syllables' have the same duration in speech, while English and French are often characterized as stressed-timed and syllable-timed respectively. Though these distinctions have been criticized (Dauer, 1983; Roach, 1982) and other classifications have been proposed (Wenk & Wioland, 1982), it is clear that these languages differ in their timing structure. Patel and Daniele (2003a) found that there are systematic differences in note durations used by French and English composers which are in accordance with measures of syllable durations in speech. It is likely that these differences are also reflected in music performance (as opposed to musical scores). Several interesting statements have been made about performances of western music by Japanese musicians. For example, the Japanese conductor Hideo Saito (1999) described a specifically Japanese interpretation of Mozart's pieces taught by music teachers in Japanese colleges of music, which seemed to mirror the Japanese language. Minao Shibata (Shibata, 1983), a well-known Japanese composer, stated that young Japanese musicians feel that the musical sensitivity of the Japanese is substantially different from that of westerners. However, these claims are mostly based on subjective impressions. Is there any empirical evidence that supports them?

Gabrielsson (1987) analyzed the timing of five pianists' performances of the first eight measures of Mozart's Piano Sonata in A-major (K331). In this study, he showed that there was a tendency for pianists to play a rhythmic pattern consisting of a dotted eighth note, a sixteenth note and an eighth note () with a shortening of the sixteenth note and a lengthening of the surrounding dotted eighth note and eighth note. However, one Japanese pianist sometimes played with the reverse tendency; that is, with a slight lengthening of the sixteenth note and a shortening of the surrounding notes. Ohgushi (2002) carried out measurements of the physical duration of intervals from the same Mozart Piano Sonata as performed by Japanese and western pianists. He reported a different tendency in the timing of the rhythmic patterns consisting of three intervals with a ratio of 3:1:2. Japanese pianists tended to produce a significantly smaller duration ratio of the first two intervals (3:1) than western pianists. This finding suggests that there are physical characteristics distinguishable within each culture with regard to the expressive timing of particular rhythms. However, to be able to generalize this observation, a more detailed and fundamental study is needed which compares performances of simple temporal patterns by musicians who have different cultural backgrounds.

Empirical research has provided considerable evidence of general rules for performing time intervals. Small integer ratios of duration are easier to process than more complex ratios; in particular, isochronous rhythm is known to be more precisely produced (Repp, Windsor, & Desain, 2002), although perfect integer ratios never appear in performance (Clarke, 1999). The so-called Short-Long (S-L) tendency, which is a slight lengthening of the second interval duration, has been observed in a performance of isochronous rhythm (Drake & Palmer, 1993; Gabrielsson, 1974; Repp, 1999d). Gabrielsson (1974) observed various deviations from exact timing in the performance of rhythmic patterns having different intervals, often determined by such factors as rhythmic structure and phrasing. Because there are many possible musical contexts, it is difficult to find a useful explanation for these deviations. In production studies that use rather simple rhythmic patterns, a tendency towards assimilation has been noted, which is termed by Fraisse (1956) as the equalization of duration differences (Sternberg, Knoll, & Zukofsky, 1982). This finding seems to be reasonably well explained by the ratio simplification hypothesis, which states that complex duration ratios tend to be reproduced as more simple duration ratios (Fraisse, 1982; Povel, 1981).

In our experiment we used rhythmic patterns consisting of two intervals with simple duration ratios of 1:1, 2:1, 3:1, 4:1 and 5:1. Reverse versions, i.e.

duration ratios from 1:2 to 1:5 were included as well. Japanese and Western musicians were compared with regard to timing in the production of these patterns. To investigate how far these differences are under the control of expressive interpretation, two conditions were used: (a) musical and (b) mechanical. Furthermore, we investigated the influence of tempo on the production of the rhythmic patterns.

Method

Twelve percussionists participated in the experiment. Six of them were professional musicians residing in the Netherlands. The other six were percussion majors (five undergraduates and one graduate student) at Kyoto City University of Arts in Japan. The Dutch and Japanese participants had an average of 21.5 years and 17.8 years of musical training respectively. This period not only reflects percussion training, but also any kind of training in western classical music.

Participants were asked to perform several rhythmic patterns from a score written in common music notation. Each rhythmic pattern consisted of two intervals whose duration ratios were 1:1, 1:2, 2:1, 1:3, 3:1, 1:4, 4:1, 1:5, and 5:1. To make sure that measurements reflect the participants' intention of performing nine different rhythmic categories, the nominal ratios of the rhythmic patterns were given as well, as an aid to a precise understanding of the patterns. This is because patterns with large duration ratios, such as 1:4 and 1:5, resemble each other more with regard to duration and they require rather complex notation. Participants were instructed to perform the rhythmic patterns at three different tempi: 60, 75 and 90 beats per minute. They practiced to get an accurate sense of tempo using a metronome before each trial. Furthermore, they were requested to perform in two modes: mechanical and musical. Participants were given a different set of scores and instructions for each mode. In the mechanical mode, participants were given a score which showed one rhythmic pattern consisting of two notes, as shown in Figure 2.1a. They were instructed to repeat this pattern as accurately as possible 10 times. In the musical mode, participants were given a score with a common time signature with four bars containing 12 rhythmic patterns in the first three bars and one quarter interval in the last bar, as shown in Figure 2.2.1b. For the musical mode, participants were instructed to perform as if they were performing a short piece of music. Numbers in both scores indicated the nominal duration ratio of two intervals forming a rhythmic pattern.

Participants were instructed to perform with one hand, using a wooden stick. A piezo contact microphone encapsulated in a box covered with a thin sheet of rubber was used as the drum surface. The performances were recorded on digital audio tape (DAT) using a sampling frequency of 44.1 kHz.

Analyses

The DAT recordings were converted to audio files (AIFF, 44.1kHz) using SoundScope (version 3.0, GW Instruments). These audio files were played through a D4 drum machine while recording its musical instrument digital interface (MIDI) output in Opcode Vision (version 4.0, Opcode). Sensitivity controls were set by hand to capture all hits. The MIDI files were converted to tables of interval durations, which indicate the time interval from an onset of one interval to the onset of the following interval, using POCO (Desain & Honing, 1992). Analyses of these intervals were conducted in JMP (Version 3.2.2 & Version 4.0, SAS).

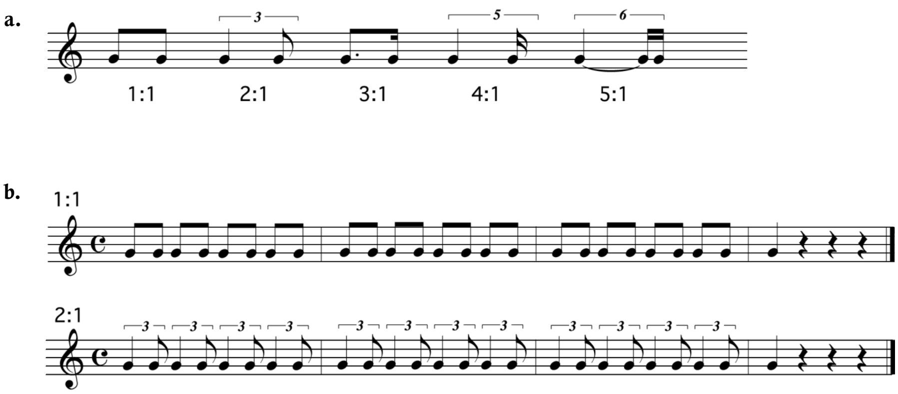



Figure 2.1 **a.** Example of the scores provided to the participant for the mechanical mode. **b.** Example of the scores provided to the participant for the musical mode.

The first and last rhythmic patterns of each trial were eliminated to avoid unstable portions. In the musical mode, the last quarter note was also eliminated. Thus the number of repetitions in one performance trial was 8 for the mechanical mode, and 10 for the musical mode. Any rhythmic pattern whose performed duration ratio of two intervals differed by more than 50 % from the instructed ratio was considered an error and excluded from the analyses. As a result, the number of trials used for analysis was 307 for Dutch and 322 for Japanese, out of a total of 324 for each participant group.

Results

The presentation of the results focuses on the effects of the mode, the tempo, the participant group, the rhythmic pattern and the temporal order of the intervals within the rhythmic patterns on the performed duration ratio. The timing of each performance was analyzed by calculating the duration ratio of intervals included in one rhythmic pattern. The performed duration ratio was calculated as the longer interval divided by the shorter interval. In the case of 1:1, the performed duration ratio was calculated as the second interval divided by the first interval. Furthermore, the proportion of deviation of performed duration ratios from the instructed ratio was computed for statistical analyses that compare rhythmic patterns having different duration ratios. Standard deviations (SDs) of duration ratios between the trials were used to study the distribution of performance timing. Performance consistency was studied by analyzing the SD of duration ratios within the trials. For the mechanical mode the SD within the trials was used to evaluate the difficulty of performance. Note that here we refer to two kinds of SDs showing different aspects of the data.

The effect of mode

Previous studies have shown that parameter variation decreases in mechanical performances (Drake & Palmer, 1993; Gabrielsson, 1987; Palmer, 1989; Seashore, 1938). Therefore the consistency of performance in the mechanical mode was expected to be greater than in the musical mode. Performances in the mechanical mode might have been expected to be closer to exact timing than in the musical mode, because of the instruction factor, but contrary to our expectations, no difference between the modes was observed for the performed duration ratio. An ANOVA on the averaged proportion of deviations of duration ratios by mode (2) and rhythmic pattern (9) showed no significant main effect of mode, (F (1, 611)=1.55, p= .21), and no significant interactions, although there was a strong main effect of rhythmic pattern, (F (8, 611)=24.93, p< .01). We also studied the effect of the metric position on the performance timing for the musical mode performances. An ANOVA on the averaged proportion of deviations of duration ratios by metric position (4) and rhythmic pattern (9) showed no significant effect of metric position, (F (3, 1244)=0.27, p= .84), a strong main effect of rhythmic pattern, (F (8, 1244)=58.17, p< .01), and no significant interactions. These show that the performed duration ratios in the case of the musical mode were stable regardless of metric position. The results obtained here also may suggest that the tendencies found in this study were not a result of the expressive interpretation of the participant playing musically, as the same trends show up in the mechanical playing mode.

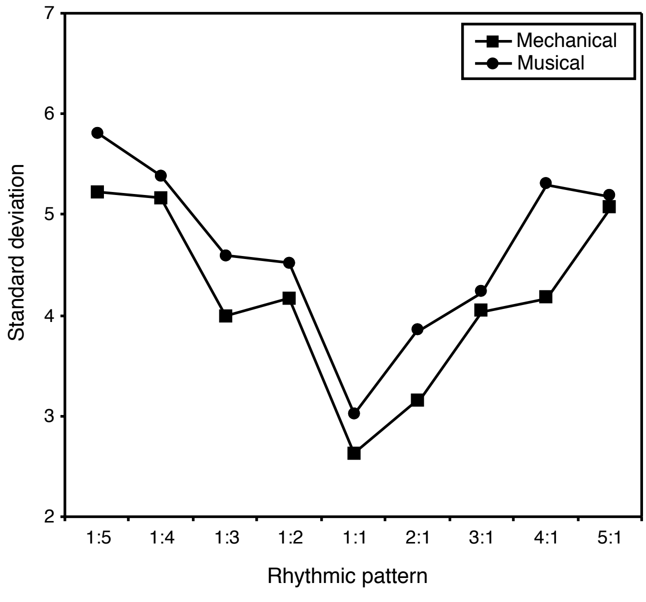


Figure 2.2 Average standard deviations of duration ratios within trials for the nine rhythmic patterns and two modes.

Although the mean performance timing did not change with instruction, less variation in the timing within each trial was observed in the mechanical mode. Figure 2.2 shows the averaged SD of duration ratios within the trials. The vertical axis shows the SD, and the horizontal axis shows the rhythmic patterns. We see here that participants perform the trials more consistently in the mechanical mode than in the musical mode. An ANOVA on the averaged SD within the trials by mode (2) and rhythmic pattern (9) showed an obvious significant main effect of mode, (F (1, 611)= 11.81, $p< .01$); there was also a strong main effect of the rhythmic pattern, (F (8, 611)=17.02, $p< .01$). There was no significant interaction between them. We considered the SD within the trials in the mechanical mode to be the standard of difficulty of the task.

The effect of tempo

We examined the effect of instructed tempo on performed tempo averaged over each performance data set. First, we studied to what extent a constant tempo is maintained. The beat durations were measured as the total length of the first and second interval's duration. Performed tempo averaged over each performance data set can be calculated from the beat duration of each repetition of the rhythmic patterns. Data distributions of performed beat duration on three instructed tempi are plotted in Figure 2.3. The vertical axis indicates the number of trials and the horizontal axis indicates performed tempo. It shows that participants performed rhythmic patterns with a wide range of tempi. The mean value of performed tempo on each tempo condition and their SD were as follows: Tempo 60 (mean=62.21, SD=3.97), Tempo 75 (mean=73.76, SD=4.24), Tempo 90 (mean=88.04 SD=4.72). An ANOVA on the proportion of deviation of performed tempo by instructed tempo (3) revealed a strong significant effect of the instructed tempo, (F (2, 626), $p< .01$). Tukey-Krummer's multiple comparisons indicated that the proportion of deviation of performed tempo was significantly slower in the cases of Tempo 75 and Tempo 90 as compared to Tempo 60 (LSD, $p< .05$). A similar tendency to play increasingly slower than instructed when attempting faster tempi was shown in Repp, et al. (2002). There was a slight tendency to perform rhythmic patterns with more tempo variation in faster tempi as shown in Figure 2.3.

We also studied the effects of instructed tempo on the interval duration ratios. Clarke (1985; Clarke, 1987) presented evidence that the interval duration ratios of piano performances are not the same in different tempi. Desain and Honing (1994) also found that performance timing is not invariant in different tempi. However, Repp (1994) found invariance of the timing profile of the performance in different tempi. Thus there is conflicting evidence concerning the effects of tempo on these performance characteristics.

Figure 2.4 presents mean values of the performed duration ratios and the SDs between the trials by three tempi, showing that the duration ratios did not vary with tempo. An ANOVA on the averaged proportion of deviations of duration ratio by tempo (3) and rhythmic pattern (9) showed no significant main effect of tempo ($F(2, 602)=0.75$, $p= .47$), and no significant interactions, although there was a strong main effect of the rhythmic pattern ($F(8, 602)=24.67$, $p< .01$).

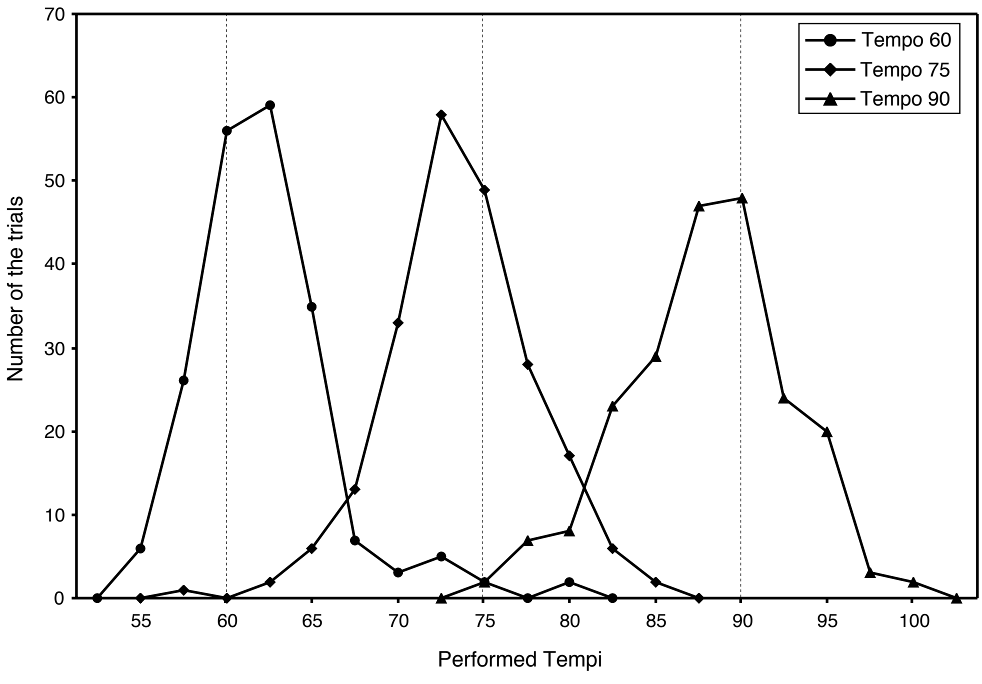


Figure 2.3 Distribution of performed tempi for the three instructed tempi.

The result corresponds to the finding of one of the previous studies, which is that the relative duration ratio of two intervals is stable for different tempi (Repp et al., 2002). However, Repp et al. also found that the duration ratios varied with tempo in the case of three-interval rhythmic patterns in the same study. Even in the case of two different intervals (2:1), in jazz performance, there is evidence that the performed ratio varies largely with tempo transposition in spite of the variety of performance style (Friberg & Sundström, 2002). These conflicting results may have come about because music was analyzed that exhibited varying features, such as complexity of rhythm, length of notes and genre (classical, jazz, rhythms without musical context). In other words, the effect of tempo on performance seems to be related so intimately to these features that it is hard to state a general rule that covers all cases.

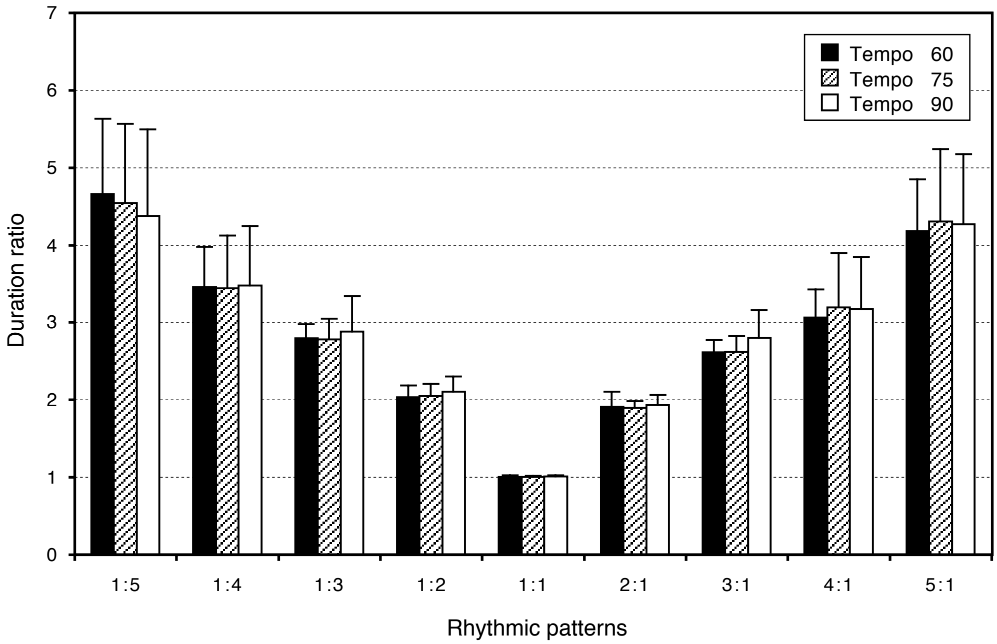


Figure 2.4 Averaged duration ratio for the nine rhythmic patterns and three tempi.

The effect of cultural background

We expected to find differences in performance timing between the western (Dutch) and Japanese participants in the simple rhythmic patterns. According to Gabrielsson (1987) and Ohgushi (2002), most western pianists have a tendency to play a 3:1 pattern with a larger ratio than Japanese pianists. Therefore Dutch players in this study were expected to perform this rhythmic pattern with two different intervals with a larger ratio than the Japanese participants.

Figure 2.5 shows mean values of the performed duration ratios and the SDs between the trials by two participant groups. The vertical axis indicates the value of the obtained duration ratio and the horizontal axis indicates the kinds of rhythmic pattern. The SD between the trials shows that Dutch participants seem to perform all rhythmic patterns with more diverse timing than Japanese participants. There were some duration ratios performed by Dutch participants above the instructed ratio in every rhythmic pattern, while the values performed by Japanese participants were seldom above the instructed ratio for large ratios such as 1:4, 4:1, 1:5 and 5:1.

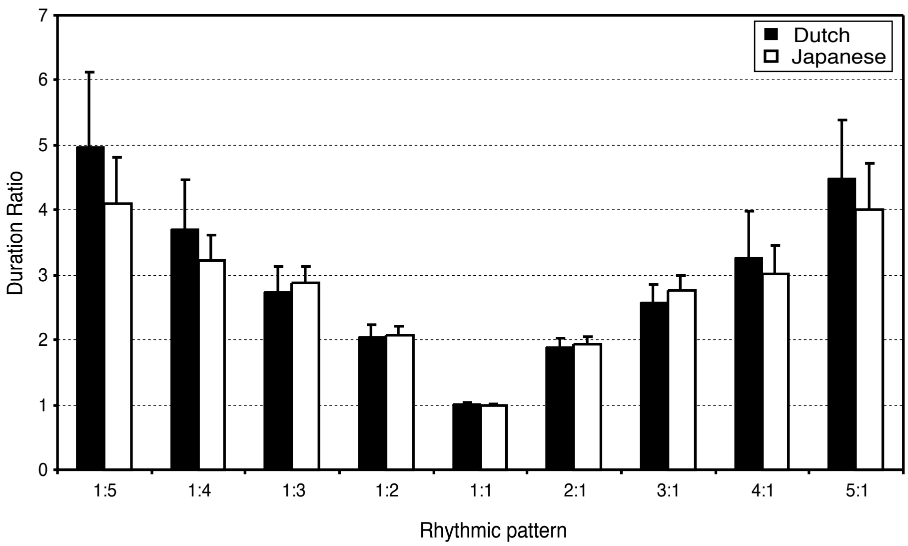


Figure 2.5 Average performed duration ratios for the nine rhythmic patterns and two participant groups.

An ANOVA on the averaged proportion of deviations of duration ratio by participant group (2) and rhythmic pattern (9) indicated that averaged deviations of Dutch participants were smaller than those of Japanese participants, ($F(1, 611)=12.28$, $p< .01$). The averaged proportion of deviations of duration ratio also varied with rhythmic patterns, ($F(8, 611)=27.44$, $p< .01$). There was a significant interaction, ($F(8, 611)=8.04$, $p< .01$), indicating the degree of deviation for some rhythmic patterns was significantly different between the two participant groups. Tukey-Krummer's multiple comparisons indicated that the averaged duration ratios were significantly smaller for the Japanese than for the Dutch in the case of 1:4 and 1:5 (LSD, $p< .05$).

Although the statistical analyses revealed a significant difference between the participant groups in some rhythmic patterns, the tendency we had expected was not systematically present in the results. However, there did seem to be a trend which distinguished the Japanese participants performing the 4:1, 5:1, 1:4 and 1:5 patterns with a smaller duration ratio than the ratio given by the scores.

The effect of rhythmic patterns

On the basis of findings from previous studies that investigated timing in music, we expected the rhythmic patterns with a smaller ratio to be performed more consistently and more closely to the value specified in the notation. In particular, the isochronous rhythmic pattern (1:1) was expected to be performed with the S-L tendency and timing close to the instruction. We also expected rhythmic patterns having two different intervals to deviate towards assimilation.

As shown in Figure 2.5, the 1:1 pattern was the one most precisely produced. The averaged proportion of deviation was surprisingly small (Dutch: 0.78%, Japanese: 0.15 %). As expected, a slight S-L tendency was observed in the performance of this pattern. Most of the durations of two different intervals showed assimilation. The larger distributions were observed in the larger duration ratios, as represented by the SD between the trials shown in Figure 2.5. Although the average duration ratios of 1:4 and 1:5 patterns performed by Dutch participants were closer to the exact ratio, there was a great variety of timings in the performances of the Dutch participants.

However, the relation between degree of distortion and the size of duration ratio was not proportional. The rhythmic pattern performed with the largest deviation was the 4:1 pattern (Dutch: 18.18%, Japanese: 24.77%) and there was also a large deviation in the case of the 1:4 pattern in the Japanese participant group. Furthermore, the averaged SD within the trial in the case of 1:5 and 5:1 patterns was not always higher than that of the 1:4 and 4:1 patterns (see Figure 2.2). Intuitively, the 1:4 and 4:1 patterns are special and awkward compared to other patterns because participants needed to divide a one-beat duration into five equal parts as quintuplets, which is rarely used in western music. One of the possible interpretations of this result is in line with the theory of the coding of temporal patterns (Povel, 1981; Povel & Essens, 1985); since the 1:4 and 4:1 patterns cannot be broke down into hierarchical small whole-number ratios, it is likely that they exhibit more deviations in performance timing than the other patterns. Thus, we could say that the degree of deviation in the performance timing seems to be not only determined by the size of duration ratio but also by its complexity in terms of subdivision. These results match the tendencies found in previous studies. Simple duration ratios such as 1:1, 1:2 and 2:1 were produced more consistently within trials. Performed duration ratios of these patterns were also consistent between the trials. Observed assimilation of two different intervals in relatively complex duration ratios such as 1:4, 4:1, 1:5 and 5:1 can be interpreted as an indication of ratio simplification.

The effect of temporal orders of the rhythmic patterns

Next, we studied the influence of the temporal order of the rhythmic patterns. The nine rhythmic patterns can be divided into three classes according to the temporal order of the nominal duration of the intervals:

- Long-Short: The first interval's duration is longer than the second (2:1,3:1,4:1,5:1)
- Short-Long: The second interval's duration is longer than the first (1:2,1:3,1:4,1:5)
- Even: The first interval's duration is the same as the second (1:1).

The long-short patterns were expected to be more familiar to the participants because successive repetition of long-short pattern is more common in western music than short-long patterns. Repp et al. (2002) found that the 1:2 pattern is more difficult to perform than the 2:1 pattern as participants had to repeat the trials of 1:2 patterns more often than the 2:1 patterns to achieve their desired performance. Similarly, the long-short patterns may be performed more consistently than the short-long patterns. The SD within the trials of long-short patterns in mechanical mode was indeed significantly smaller than that of Short-Long patterns ($p < .05$), supporting our expectation.

Furthermore, an effect of temporal order was not only observed for the consistency of the performance but also for the performed duration ratio, as participants tend to perform smaller duration ratios in long-short patterns than in short-long patterns, in spite of being instructed to perform the same duration ratio (see Figure 2.5). An ANOVA on the averaged proportion of deviations of duration ratio by order (2, excluding the 1:1 pattern) and instructed duration ratio (4) revealed the significant effect of order (F (1, 551)=30.31, $p < .01$). There was no significant interaction between order and instructed duration ratio, indicating that the effect of order was the same for each instructed duration ratio.

Although no significant difference in consistency of performance between long-short and short-long pattern was observed, there was a systematic relation between the temporal order of the two intervals and the degree of deviation of the performance timing.

Discussion

Our main aim in this chapter was to pinpoint cultural differences in playing style of simple rhythmic patterns. Although we have not found very large effects, we have found various characteristics of performance timing, e.g. an independence of tempo and performance mode. One may argue that the two participant groups were not matched well enough, as they differed not only in nationality but also in the amount of musical training experience (Dutch 21.5 years, Japanese 17.8 years) and degree of professionalisation. As the length of musical training experience may not be proportional to skill in musical performance, judging the match is quite difficult. However there are some studies which suggest that professional musicians are better than amateurs or students at producing consistent timing patterns (Gabrielsson, 1987; Palmer, 1989; Repp, 1990; Sundberg, Prame, & Iwarsson, 1996). This suggests that we can determine whether the two groups are adequately matched in terms of skill by measuring within-participant variability.

The SD within the trials of mechanical performance data was applied to compare two participant groups with respect to performance stability. The SDs of duration ratios within the trial of each participant for every rhythmic pattern were averaged. The t-tests on these averaged SDs showed no significant differences between Dutch and Japanese groups, $p= .43$. Although the average number of years of musical training of Japanese participants was smaller than that of the Dutch participants, it seems that performance consistency was not significantly different between the two groups. Furthermore, one might argue that the two groups compared in this study are not clearly culturally distinct, as both groups of musicians have had training in western classical music. However, the environments in which these two groups live their daily lives are clearly different, and the languages they use and are exposed to are also grammatically and rhythmically unrelated.

For some of the large ratios (1:4 and 1:5), there were significant timing differences between the two groups, as Japanese participants performed them with smaller duration ratios than Dutch participants; the same tendency was observed for the reverse version of these patterns (4:1 and 5:1), although the effect is not significant in these cases. This tendency was not found for small ratios (1:1, 1:2, 2:1, 1:3 and 3:1). Deviations of the duration ratio from the instructed ratio were larger for large ratios. The SD between trials also increased in the large ratios. All these observations suggest that the performance timing of large ratios was more unstable than that of small ratios. This seems to be in accordance with the general principle that small integer ratios are easier to cope with. In other words, the rather simple process of performance timing is common to both participant groups, whereas cultural differences are more likely to occur in the processing of more complex rhythms.

However, rhythmic patterns in real music have a far richer structure than the rhythms used in this study. The relation between performance timing and more complex rhythmic structures has yet to be investigated, and we would expect that both common and independent cultural features of the musical performance will be found to play a role.

Chapter 3

Comparing rhythmic structure in popular music and speech

Abstract

Recently, empirical evidence of an influence of the prosody of language on composed musical rhythm has been demonstrated. In the current study, two factors, which shape rhythmic structure, were distinguished: an underlying distribution of the durations, and a tendency of sequencing note durations towards more or less contrasted alternations, as measured by the normalized Pairwise Variability Index (nPVI). The main question was which of these factors contributes more to form the different characteristics of analyzed rhythms. Rhythms in popular music by English and Japanese composers, as well as in British English and Japanese speech were examined. The analysis revealed that the underlying probability distribution of durations contribute to a significant degree as a cause of the difference in patterns. This held both for music and speech. Another question investigated was to determine whether the mother tongue has an influence on the rhythms in popular music. Rhythms by UK composers and Japanese composers showed a significant contrast as UK composers tended to use more contrasted durations successively in their rhythmic sequences than Japanese composers. This tendency is in line with that of speech rhythms in British English and Japanese. We further showed that the patterns of occurrence of short-long and long-short combination of durations in the rhythms of music and speech were different between but similar within the cultural groups. The current study adds another reference for the discussion of a cognitive association between music and speech domains.

What aspects of culture have an effect on how music is written, performed and perceived? The impact of culture on music has been a key topic in musicology for a long time. Among the various aspects of cultural heritage, much effort has been made to associate the characteristics of language with different aspects of musical behavior. For instance, Lerdahl and Jackendoff (1983) pointed out that, though very different, the organization of both music as well as language can be viewed as a hierarchical syntactic structure. In addition to such structural comparisons, many empirical studies comparing the cognitive aspects of these domains have been carried out. For instance, the way in which speech phrases are expressed through timing is found to be similar to that of musical phrases, e.g., phrase final lengthening (Honing, 2003; Sundberg & Verrillo, 1980; Todd, 1985). In fact, timing patterns seem to be so crucial for human communication that infants already make use of them as a perceptual cue for understanding the phrase structure of speech and music (Hirsh-Pasek, Nelson, Jusczyk, Cassidy, Druss, & Kennedy, 1987; Jusczyk & Krumhansl, 1993; Krumhansl & Jusczyk, 1990; Nazzi, Bertoncini, & Mehler, 1998). Furthermore, there is evidence of phenomena common to both domains at the fundamental perceptual level of timing and pitch, such as categorical perception (Clarke, 1987; Desain & Honing, 2003; Repp, 1984; Siegel & Siegel, 1977; Studdert, Liberman, Harris, & Cooper, 1970) and restoration (De Witt & Samuel, 1990; Samuel, 1981). These parallels between music and speech naturally lead to growing interest in the possible communalities in their neural basis (Besson, Faita, Peretz, Bonnel, & Requin, 1998; Patel, 1998; Patel, Peretz, Tramo, & Labreque, 1998; Peretz, Gagnon, Hebert, & Macoir, 2004a; Peretz, Radeau, & Arguin, 2004b).

As music and language are both conveyed as sequences of sounds organized in time, the temporal aspects are highly important for both domains. Wenk (1987) showed that certain musical works written by French composers tend to reflect the prosodic characteristics of the French language. More recently, Patel and Daniele (2003a) showed that music composed by French and English composers of a certain era reflected the rhythmic differences between the composers' native languages.

Patel and Daniele (2003a) applied a measure, the so-called normalized Pairwise Variability Index (nPVI), to the themes of western classical music. The nPVI was originally developed to describe rhythmic differences between languages by Ling, Grabe and Nolan (2000). Given a sequence of m intervals of duration $d_1 d_2 ... d_m$, the nPVI is defined as:

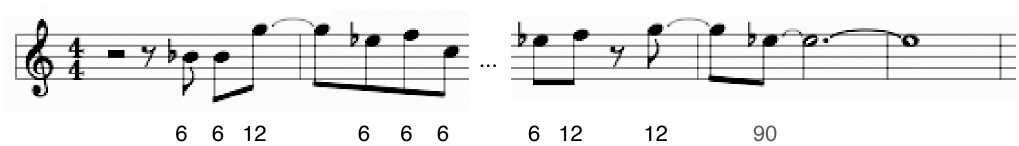
$$nPVI(d_1 d_2 ... d_m) = \frac{200}{m-1} \times \sum_{k=1}^{m-1} \frac{|d_k - d_{k+1}|}{d_k + d_{k+1}} \qquad (1).$$

For each pair of subsequent intervals, the contribution to the nPVI is their difference divided by their means. The average of these differences is taken and expressed as a percentage.[1] This index has a theoretical minimum of 0 (for isochronous sequences) and a theoretical maximum of 200 (for sequence of alternating long/short durations whereby the short durations approach 0). Different from standard deviation and variance, the nPVI is a measure of the contrast of successive durations and is sensitive to the order of the durations. This index is founded on the principle that there is usually vowel reduction (or even absence) observed in the unstressed syllable in stress-timed languages. This seems to result in greater durational variability of vowel intervals than in syllable-timed languages (Dauer, 1983, 1987; Nespor, 1990). Also syllable-timed languages exhibit a simpler syllable structure than stress-timed languages, which is supposed to contribute to the smaller durational variability. In these studies, the nPVI is applied to analyze the vocalic intervals for controlling speech rate (Grabe and Low, 2002).

A difference among stress-timed, syllable-timed, and mora-timed languages has been empirically shown using this measure (Grabe & Low, 2002; Ling et al., 2000; Ramus, 2002). In these studies, stress-timed languages (e.g., English, Dutch) showed relatively high nPVIs for vocalic intervals, whereas syllable-

[1] Although we opt for the shorter description, the original form of the formula is expressed as:

$$nPVI(d_1 d_2 ... d_m) = \frac{100}{m-1} \times \sum_{k=1}^{m-1} \left| \frac{(d_k - d_{k+1})}{(\frac{d_k + d_{k+1}}{2})} \right| \qquad (1a).$$



$$nPVI = \frac{200}{62-1} \times \left[\left|\frac{6-6}{6+6}\right| + \left|\frac{6-12}{6+12}\right| + \left|\frac{12-6}{12+6}\right| + \left|\frac{6-6}{6+6}\right| + \left|\frac{6-6}{6+6}\right| \cdots \left|\frac{6-12}{6+12}\right| + \left|\frac{12-12}{12+12}\right| \right] = 30.1$$

Figure 3.1 An example of a musical fragment. The relative duration of each note is shown below the musical score. The bar is divided into 48 units and the duration of each musical note was assigned accordingly. The calculation of the musical nPVI for this example is given in the formula.

timed languages (e.g., French, Spanish) showed a lower vocalic nPVI. The contrastiveness of the intervocalic intervals measured by rPVI (raw Pairwised Variability Index), which calculates the pure distance between successive durations without normalization, was about the same between these languages. Although the studies did not agree on the results of intervocalic rPVI in a mora-timed language (Japanese), Japanese vocalic nPVI has been shown to be similar to that of syllable-timed languages. Although there are not clear-cut categories of languages (Roach, 1982; Wenk & Wioland, 1982), these measures still seem to offer a robust distinction.

Turning to music, Patel & Daniele (2003a) showed that the nPVI values calculated from musical scores composed by native English speakers are relatively high, as compared to the musical nPVIs by native French composers. The approach has been further validated by expanding the number of languages, composers and musical eras (Huron & Ollen, 2003; Patel & Daniele, 2003b). This correlation between rhythm in language and music is very interesting, as it could be interpreted as an indication that temporal information processing mechanisms are shared by speech and music. In these studies, the notated, symbolic duration of musical notes was used. As the nPVI considers proportional differences in successive durations (sequences such as 1-3-2 and 10-30-20 yield the same nPVI value), an arbitrary unit can be chosen for the duration of the bar and each musical note can be expressed in that unit. Figure

3.1 shows an example of a short musical score and the duration coding. We used 48 as the unit.

In the current study, we investigated issues which still need to be clarified in regards to the musical nPVI approach. The first issue concerns the source of nPVI differences. A high nPVI has always been interpreted as a preference to order durations in a contrasting sequential manner. However, it could also be caused by, for example, the more extreme long and short durations and a random order. Thus, the differences shown in the nPVI studies may be caused by a) a different underlying distribution of the note durations used in the compositions or b) a different preference for sequencing note durations. To distinguish these hypotheses, we examined duration histograms and the value of the expected index if a random ordering were to be used. Distributions of the duration occurring in the corpora of popular music written by native English and Japanese speakers were compared to check if the underlying characteristics of the data sets were already different. Furthermore, a measure was developed which characterizes the contrast arising from the underlying distribution itself. The expected nPVI (E-nPVI) assuming random sequence of the durations was calculated and was compared with the real nPVI to see if there was an effect of sequencing durations. If the differences only stem from a different choice of durations and not from a different tendency to order the note durations, the E-nPVIs should be in the same value range as the nPVIs. On the other hand, if different tendencies in ordering note durations determine the musical nPVI differences, the E-nPVIs should be similar between cultures while the nPVI is not.

Though the nPVI is better at capturing characteristics of the order of the durations compared to summary statistics such as standard deviations and averages, it is insensitive to some ordering differences, such as the time direction of sequencing note durations, for example, distinction between 1-3 and 3-1. Such short-long and long-short patterns yield an identical nPVI value, while they imply very different rhythmic structures. According to Longuet-Higgins and Lee (1984), 1-3 evokes a strong syncopation feeling while 3-1 does not. Difference in occurrences of these short-long and long-short combinations may contribute to the formation of different impressions of music. It has been shown that long-short combinations occur more than short-long combinations, when the ratios of note pairs that span a metric unit together (musical bar and beat) are counted. This occurrence of ratios in music scores has been further shown to be a good candidate measure of a prior knowledge of rhythm perception in music (see Sadakata, Desain & Honing, 2006). However, such a strong effect of hierarchical association of durations does not exist in speech rhythm except for in verse and poetry. Thus, it is interesting to align music and speech by ignoring such hierarchical organization of music and taking only serial ordering into account. To this end, we separated the occurrence of the long-short and short-long patterns and checked how they contribute to the nPVI.

Note that the cause of the nPVI differences and the effects of time direction have not been examined on speech data yet. Thus, the same method is applied to the corpora of British English and Japanese speech (Ramus, Nespor, & Mehler, 1999).

Another issue considered in the current paper is to broaden the application of the musical nPVI method. Although the link between rhythm in music and speech of so-called syllable-timed and stress-timed language has been broadly investigated, so-called mora-timed languages, such as Japanese and Finish, have not yet been investigated. This is probably due to there being very few music scores written by native speaker of other language than stress- or syllable-timed language written in the manner of western classical music from the era on which Patel and Daniele (2003a) conducted their study (17 – 20 century). However, the nPVI can be applied with relative ease to Japanese music of that has appeared in a later periods. For example, Japanese popular music is ideal in this respect because it is much more comparable with its counterpart in English as the instruments are often the same and the western tonal musical structure is mostly used. Furthermore, although traditionally it has been largely neglected by music cognition researchers (but see, e.g.,Altenmüller, Schurmann, Lim, & Parlitz, 2002; Drake & El Heni, 2003), popular music is an important domain because it has a much larger audience than any other genre and it has a great influence on people's everyday life.

For speech, Japanese and French have virtually identical nPVI values, which are lower than that of British English. In the case of French, this speech value is reflected in a lower value for French music. Thus the hypothesis is that if rhythmic structure in speech has an influence on that in music. We predict a lower nPVI value for Japanese popular music as compared to English popular music.

In the context of investigating popular music, both the mother tongue of the composer and the language of the lyrics can potentially affect the musical rhythm. It is difficult to separate out these two sources completely, because not many English-speaking composers produce music with Japanese lyrics. However, the converse is more common. A subset of Japanese popular music that has English lyrics was included in the studies to consider the effect of lyrics on musical rhythm. If the rhythm of lyrics is contributing to the formation of the rhythmic structure in music, a higher nPVI is predicted for English lyrics written by Japanese composers as compared to those with Japanese lyrics.

Materials

Music

English and Japanese popular music written by English and Japanese native speakers was compared, namely set E (native English composers) and set J (native Japanese composers). Within set E, the songs from UK and US are distinguished as E_{UK} and E_{US}, respectively. The subset of Japanese popular songs that have English lyrics is shown as J_E as contrasted with J_J of which lyrics are in Japanese.

The E set was collected from a MIDI (Musical Instrument Digital Interface) Database, Pop (n.d.). MIDI is a standardized protocol for communication for electronic music devices and computers. This database contains a large amount of MIDI files from various genres of music. The files from the Pop category were retrieved (one song per artist), and only songs by native UK or US English speakers were included. They were hit songs from 1980 to 2000. As a result, 82 MIDI files were analyzed (13 files for the E_{UK} set and 69 files for the E_{US} set). The melody track from the MIDI file was extracted, and files containing expressive timing (5 files) were carefully quantized to yield the score timing. The procedure was conducted by the first (Japanese) author, and revaluated by a third researcher (Dutch), yielding 96% inter-rater agreement. The J set consisted of 100 MIDI files of popular music from the RWC Music Database (Goto, Hashiguchi, Nishimura & Oka, 2002). The 20 MIDI files of the RWC Music Database were with English lyrics. Thus, the J_J set consists of 80 files and the J_E set consists of 20 files, respectively.

Additionally the refrains, which repeatedly appear in the melody, were analyzed independently. The refrains were found in 168 MIDI files (E_{UK} = 11, E_{US} = 60, J_J = 77, J_E = 20).

Speech

Speech data from Ramus et al. (1999) was used for analysis. We used the English and Japanese data sets from the corpus in this study. Each language set consists of four speakers pronouncing five sentences, which made the data set 20 utterances per language. Vocalic and consonantal intervals were identified in the utterances. Vocalic intervals were analyzed, because the significant difference in successive vocalic interval between English and Japanese was observed using nPVI, while consonantal intervals were treated with a different index (rPVI).

Apparatus

POCO (Honing, 1990) was used for the segmentation of musical materials and the calculation of musical nPVI. Matlab 7.0 (The Math Works) was used for the speech data calculation. Statistical analyses were conducted by JMP ver. 5.0.1 (SAS).

Method

Musical nPVI

The musical nPVI was calculated from the durations of notes in the MIDI files. Only one track of the file (the melody) was considered. The inter onset intervals of the melody were taken as the music durations. Musical phrase boundaries were solely determined by the inter onset intervals. Durations longer than one bar were omitted and regarded as the end of a rhythmic phrase and no words of lyrics were segmented. Rhythmic phrases including less than 12 successive notes were excluded in accordance with previous studies (see Figure 3.1). The melodies consisted of 461 musical phrases for set E (E_{UK} = 99, E_{US} = 362), 667 phrases for set J (J_J = 510, J_E = 157). Also, the refrains consisted of 70 musical phrases for E (E_{UK} = 10, E_{US} = 60) and 105 for J (J_J = 84, J_E = 21), respectively. One bar was divided into 48 time units and the note durations proportional to one bar were represented using these time units; e.g. a sixteenth note in time 4/4 signature is was represented as 3, and in 3/4 and 6/8 signature was represented as 4. These note durations were substituted in the place of vocalic durations in formula (1).

Expected nPVI

As one musical bar was divided into 48 units, the frequency of occurrence of 48 different durations was obtained for MIDI data. Expressed as a discrete probability density, for each interval *k*, with ($1 \leq k \leq 48$), it specifies the probability of the occurrence of that interval *p(k)*. The match of the shape of the distributions could be examined using a two-sample Kolmogorov-Smirnov test. However, this test does not provide the size and direction of the difference. As this distribution exhibits substantial internal structure, e.g., it has several peaks at certain durations and has zeros in between, there is little point in comparing the standard deviation or other summary statistics. Our interest was not to see the size and the direction of the differences in distributions but rather to see how much of the difference in distribution contributes to form the different nPVIs. In fact, the nPVI itself provides a way to ascertain whether the distribution may be the source of different nPVI values. From the distributions of durations, the expected value of the nPVI, assuming a fully random order of the notes, can be calculated. For this, the probability of any combination of two subsequent durations (the product of their individual probabilities) and the contribution of such a pair to the random nPVI was calculated. Summing over every possible pair of intervals yields an expected value of the nPVI for each given distribution:

$$E(nPVI) = 200 * \sum_{h=1}^{48} \sum_{i=1}^{48} \frac{|h-i|}{h+i} p(h)p(i) \qquad (2).$$

Results

Musical nPVI

First, the musical nPVI of set E and J were compared. Figure 3.2 presents the means and standard errors for set E and J, and for the subsets. Contrary to the hypothesis, statistical analysis showed no significant difference between the groups (Mann-Whitney U-test, U= 150092, p=.497). However, much larger nPVI difference between the cultures were observed when the nPVI from only the refrain part were compared (Mann-Whitney U-test, U=3112, p=.086). The nPVIs for the refrain parts are presented in Table 3.1.

Further analyses were conducted for the subsets. The speech nPVI difference has been shown for the British English (Grabe & Low, 2002; Ramus et al., 1999). We replicated the calculation of the speech nPVIs of the vocalic intervals from the data by Ramus et al. (1999) and the effect was confirmed, as the speech nPVI for British English was significantly higher than that for Japanese (Mann-Whitney U-test, U=41, p<.001). The mean nPVIs of the speech data can be found in Table1.2. To align our musical data with the speech data, we compared the musical nPVI of the EUK and JJ sets. Difference between the groups was highly significant (Mann-Whitney U-test, U=19964, p<.001). The refrain nPVIs also showed significant differences (Mann-Whitney U-test, U=217, p<.05).

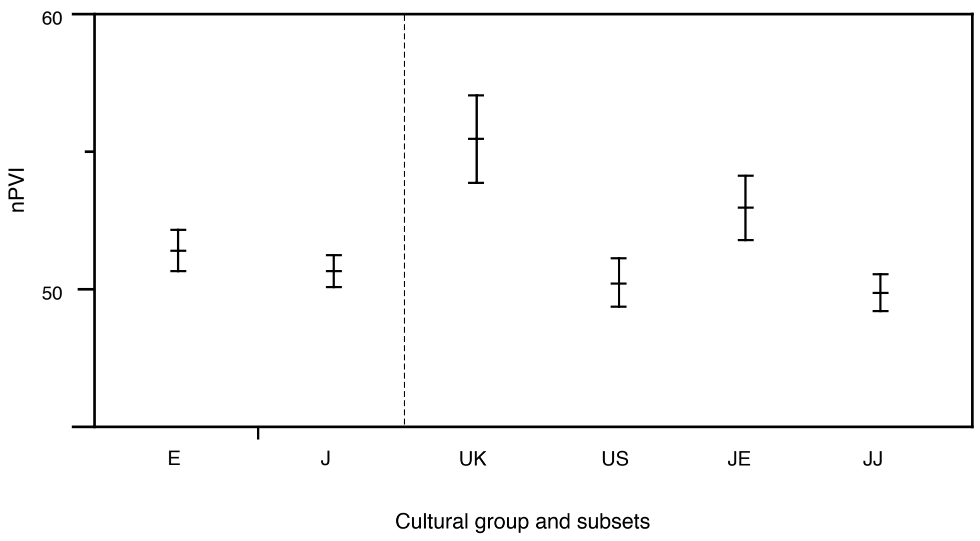


Figure 3.2 The mean and the standard errors of the musical nPVI points for rhythms written by English speaking composers (E) and by Japanese speaking composers (J). The subsets of these groups are shown in as E_{UK}, E_{US}, J_J, and J_E, respectively.

Table 3.1 The nPVI calculated from the refrains.

nPVI for refrains	
E	**54.7**
E_{UK}	60.3
E_{US}	53.7
J	**49.7**
J_J	50.6
J_E	49.5

The influence of language of the lyrics was investigated by comparing subsets of the J set; set J_J and J_E. Both corpora were composed by Japanese composers but one with Japanese lyrics while the other with English lyrics. A significant contrast between J_J and J_E was found (Mann-Whitney U-test, U=35640, p<.05), with higher nPVI values for J_E than that for J_J. Thus, with regard to nPVI, a significant effect of the language of the lyrics on the way Japanese composers making rhythmic sequences was found. Interestingly, the refrain nPVI of the J_E set was not high at all (Mann-Whitney U-test, U=821, p=.63).

The analyses in this section confirmed that the durational variability in music analyzed here showed the expected tendency, to a certain extent. Also, the language of the lyrics is likely to have an influence on how musical rhythms are organized, as J_E showed a higher nPVI value than that of J_J. The interesting findings of the refrain nPVIs will be discussed later.

Distributions vs. ordering

Having found some significant differences in the nPVIs, it is now of interest to look for the source of these differences. The frequency of occurrence of the 48 possible durations was counted. A similar pattern among data sets emerged (a sample of the histogram can be found in Figure 3.3 as *p(h)* and *p(i)*). In general, the longer the duration of a note, the less it occurred. There was also a tendency that multiples of the duration of 3, thus multiples of a 16th note in 4/4 meter, occurred more often than the other durations. The total proportion of occurrences of these musical durations was high, on average 90 %. This indicates that the occurrence of different notes seems to be highly constrained by the musical structure, as the notes in which durations were equivalent to metrically important units occurred more in all the data sets. However, a Kolmogorov-Smirnov two-sample test revealed that the distributions from set E_{UK} and J_J, and set J_J and J_E were significantly different ($p<.01$), respectively. Thus, although the distributions globally exhibit similarities, still their local characteristics were different.

The nPVI difference may indeed be caused by different underlying distributions of note durations. Yet we cannot tell how much of this difference in the distribution is contributing to the differences in the nPVI. We calculated the E-nPVI to examine the effect of the distribution and the ordering. The different distribution is likely to cause the difference in the nPVI values when the E-nPVIs are equivalent to the nPVIs. Conversely, there must be an effect of ordering that contributes to the nPVI when the E-nPVIs are different from the nPVIs.

Figure 3.3 illustrates the process of calculating the E-nPVI and the nPVI. The left column shows the random order simulation, while the right column shows the observed data. The *p(h)* and *p(i)* show the histograms of the first and second interval, respectively. The scale of the note durations used for the figure is from 1 to 24, because most of the observations concentrated in this range. Panel *a* predicts the probability of the occurrence of the combination of note durations by multiplying the two histograms, *p(h)p(i)*, as if they are independent. The true joint probability of the occurrence of note durations, *p(h,i)* shown in the panel *d*, presents the observed occurrence of duration combinations. Diagonal lines indicate the combinations in which note durations are identical. The darker the gray area, the higher the proportion of occurrence of the combination of the two intervals. The middle panels, *b* and *e*, present the abstract nPVI contribution of a specific interval pair *(h,i)*. Higher nPVIs are shown as darker shades. The bottom panel illustrates the products of the first row and the second row, which is the nPVI contribution of the occurred note duration combinations. Summing the gray values within this panel corresponds to E-nPVI (left column) and nPVI (right column).

The average nPVIs, the means of the nPVIs calculated per phrase, and E-nPVIs that were calculated per data set, are presented in Table1.2. The E-nPVI

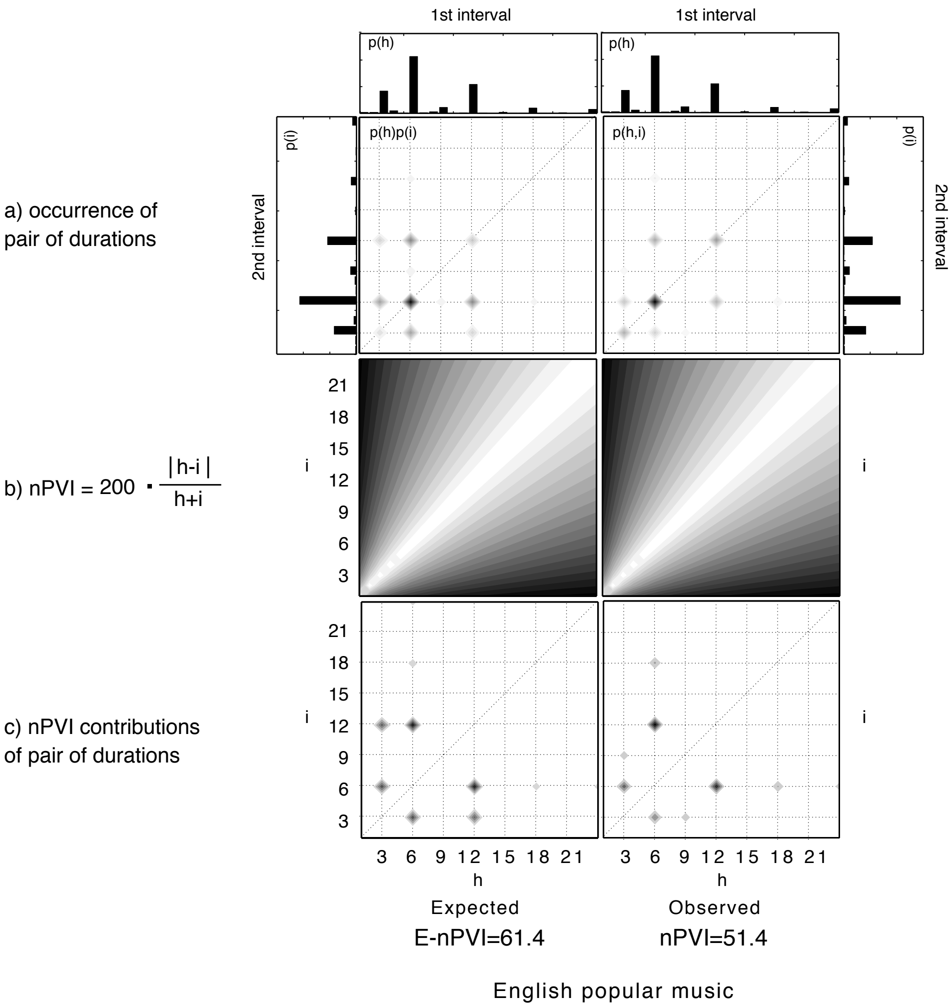


Figure 3.3 The process of calculating the E-nPVI and the musical nPVI. The left column shows the E-nPVI and the right column show the musical nPVI. Histograms *p(h)* and *p(i)* present the distribution of the occurrence of the first duration and the second duration. The left top panel shows a probability of occurrence of a combination of note durations hypothesized as independent, *p(h)p(i)*. The right top panel shows their joint probability, *p(h,i)*. The middle panels present the nPVI factor. The bottom panels present the nPVI contribution as multiplication of the probability and the nPVI factor. Summing over the surface gives the nPVI and E-nPVI value as listed below.

Table 3.2 The means of the nPVI and the E-nPVI.

	nPVI	E-nPVI
E	**51.4**	**61.4**
E_{UK}	55.5	64.1
E_{US}	50.3	60.5
J	**50.6**	**57.7**
J_J	49.9	56.7
J_E	53.0	60.5

values were always higher than the nPVIs, suggesting that the order caused an effect contributing to form the nPVIs: Composers preferred to use less contrasted durational sequences than the chance level that is predicted from the distributions. However, more importantly, the same amount of difference as in nPVI was found in E-nPVI values. This indicates that the underlying distributions contribute to a large extent to form the difference in nPVI values.

Because the nPVI contribution of a pair of intervals *(h,i)* is uniquely determined by its ratio $r=(h/i)$, we can reduce the dimensionality of Figure 3.3. We redefine the nPVI as

$$nPVI(d_1,d_2...d_m) = \frac{200}{m-1} \times \sum_{k+1}^{m-1} \frac{|r_k - 1|}{r_k + 1} \text{ with } r_k = \frac{d_k}{d_{k+1}} \qquad (3).$$

We can calculate the PVI from a probability distribution of ratios: this is illustrated in Figure 3.4. The top panels show the probability distribution of ratios, where a ratio r smaller than 1 represents a short-long combination (SL), larger than 1 represents a long-short combination (LS), 1 represents an equal durational ratio. The middle panel shows the nPVI contribution of a specific ratio, and the bottom panels show the multiplication of the probability distribution of the durational pairs and nPVI, presenting the actual nPVI contributions of each durational pair. Summing over these histograms gives the nPVI measure.

Now we turn to the effect of the time direction of sequencing note durations. The occurrence of the long-short and short-long durational ratio was

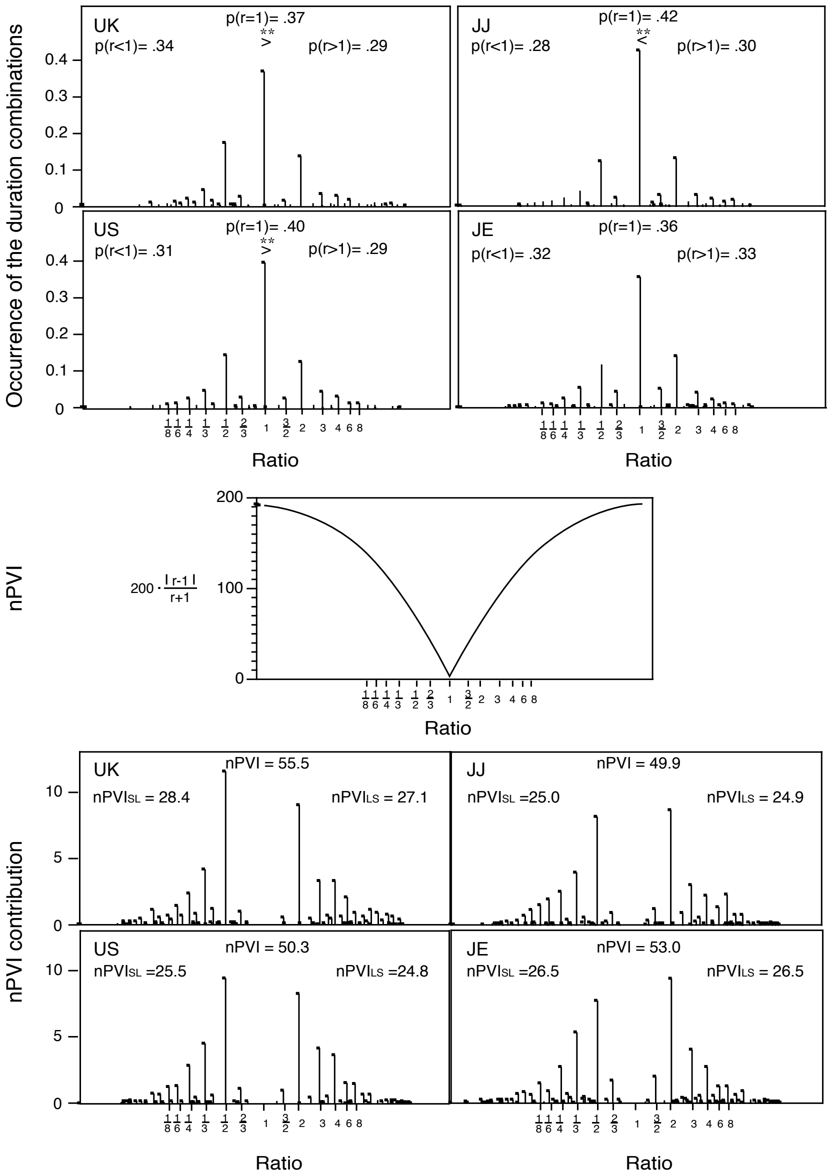


Figure 3.4 The analysis of the occurrence of short-long (SL) and long-short (LS) durational combinations for each data subset. The top four graphs indicate the occurrence of the duration ratios, the middle graph shows the nPVI factor, and the bottom four graphs indicate the nPVI contribution of the duration ratios. The total proportions and the nPVI contribution of the occurrence of SL, LS and equal durational combinations are indicated in the graphs.

counted for each data subset. The bottom four figures show the multiplication of the occurrences of the durational combinations and nPVI, presenting the nPVI contributions. The nPVI contribution of SL, LS patterns are shown as $nPVI_{SL}$ and $nPVI_{LS}$, with the sum of them corresponding to the nPVI that was calculated from the overall frequency distribution of durations for each data set. Contrary to the ratio counts within the metrical unit which showed asymmetry in occurrence between LS and SL (Sadakata et al., 2006), the proportion of the occurrence of the LS and SL were similar. In fact, SL patterns occurred significantly more than LS patterns for the E_{UK} ($Z=4.00$, $p<.01$) and E_{US} ($Z=4.01$, $p<.01$) sets, while reverse was true for J_J sets ($Z=3.73$, $p<.01$). With regard to the nPVI contribution, the ratio 1/2 contributes more than its counterpart, 2, for the E_{UK} and E_{US} sets, while it was not the case for the J_J and J_E sets.

Speech data analysis

As reported earlier, statistical test reconfirmed the significantly higher nPVI value for British English speech than for Japanese speech. It is of interest to see how much of the underlying distributions and the preference for ordering sequences contribute to this result. Figure 3.5 shows the nPVI contributions as in the same manner as in Figure 3.3, but now in a continuous scale (the nPVI panel is omitted to avoid redundancy). The *p(h)* and *p(i)* are the distributions of the first and the second durations. They are smooth, as they are obtained by Parzen's method (Parzen, 1962). The left column corresponds to the expected nPVIs from the given distributions assuming independence and the right column corresponds to the observed data. The observed nPVI and the E-nPVI are shown in the bottom of each column.

A significant difference of the duration distributions between data sets was confirmed by Kolmogorov-Smirnov two-sample test ($p<.01$). Table 3.3 indicates that the E-nPVI calculated from these duration distributions showed the parallel with the nPVI in general. The E-nPVI for British English speech was

Table 3.3 The means of the nPVI and the E-nPVI.

	nPVI	E-nPVI
English	63.8	56.2
Japanese	45.8	46.9

considerably higher than that for Japanese speech. This means that, as in music data, different underlying distributions largely determine the difference in the nPVI. The Japanese E-nPVI was slightly higher than Japanese nPVI, implying that there was a tendency in observed data to prefer less contrasted durational sequences. In contrast, the English E-nPVI was lower than the actual data for British English data, suggesting that more contrasted durational sequences were used in the data sets. Indeed, it can be seen in the panel c of English speech plot that distribution was bimodal, with peaks of the occurrence of the pairs on both sides of the diagonal line, which means that there were a large proportion of uneven pairs. In contrast, the distribution for Japanese speech was unimodal with a peak on the diagonal, which indicates more occurrences of even pairs.

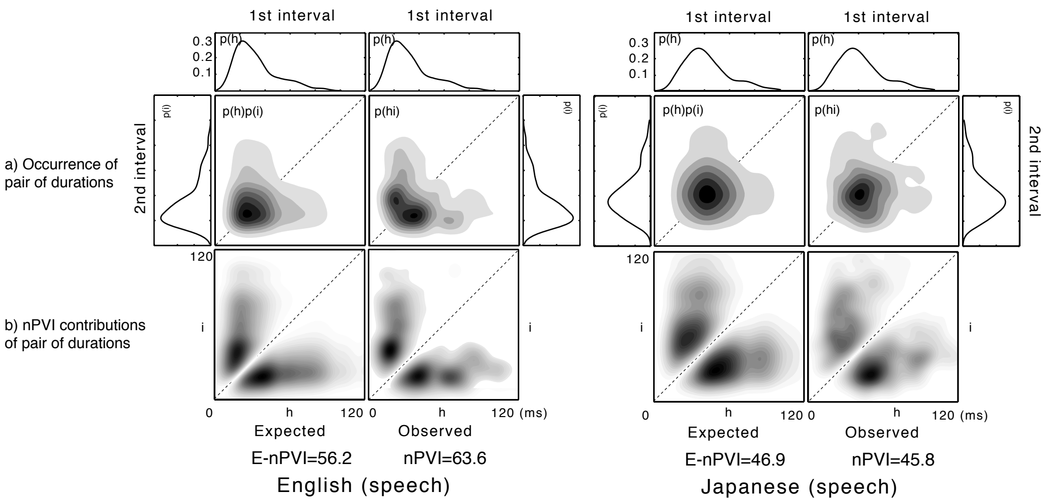


Figure 3.5 The process of calculating the E-nPVI and the nPVI on a continuous scale for the speech data by Ramus et al.(1999). The left column shows the E-nPVI and the right column show the speech (vocalic) nPVI. Histograms *p(h)* and *p(i)* present the distribution of the occurrence of the first duration and the second duration. The left top panel shows a probability of occurrence of a hypothesized independent combination of note durations, *p(h)p(i)*. The right top panel shows their true joint probability, *p(h,i)*. The bottom panels present the nPVI contribution as a multiplication of the probability and the nPVI factor. Summing over this surface gives the nPVI and E-nPVI value as listed below.

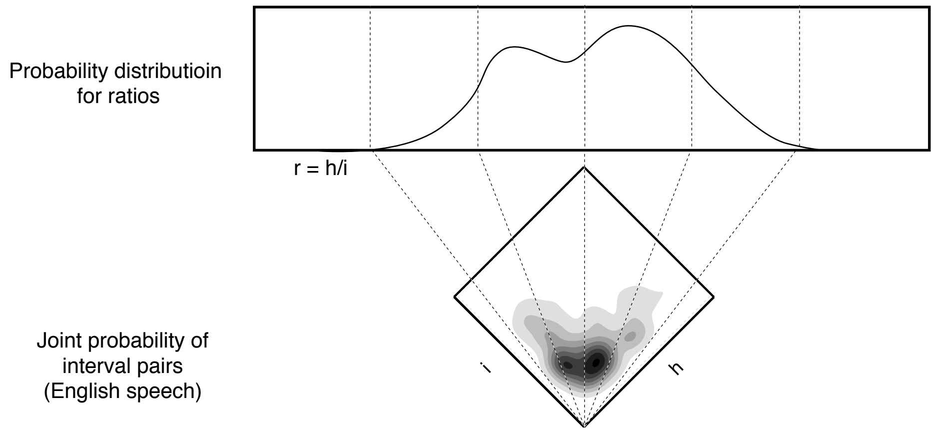


Figure 3.6 The probability distribution for ratios as projection of the joint probability of interval pairs.

As for music, we turn to a description in terms of ratios to be able to simplify the full account of the joint probability density that was given in Figure 3.5. The probability distribution for ratios can be derived from the full joint probability of interval pairs by projection as illustrated in Figure 3.6.

Now we get a clear picture of the possible asymmetry in the pattern of sequencing durations. Figure 3.7 shows the proportions of the occurrence of duration combinations, nPVI, and the nPVI contributions as the same manner as in Figure 3.4, but on a continuous scale. It is interesting to see in the top panels that English data has bimodal distribution, with peaks for both SL and LS combination, while Japanese data exhibits a single peak at the ratio of around equal length combination. Thus, English vocalic interval pairs tended more towards the extreme, which is of course reflected in the high nPVI for English. Although a difference in occurrence of SL and LS patterns did not reach the significance level, the nPVI contribution of SL is more pronounced in English speech data.

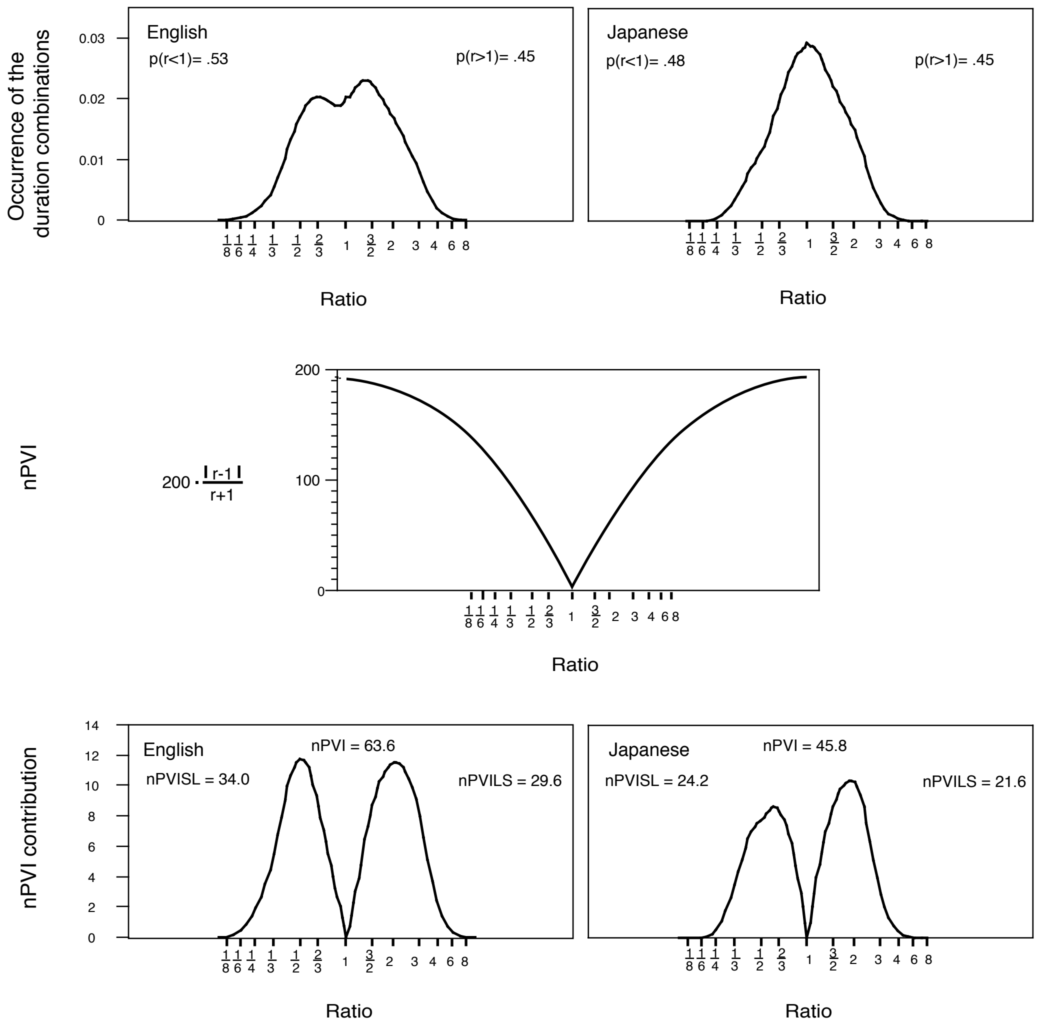


Figure 3.7 The analysis of the occurrence of short-long (SL) and long-short (LS) durational combinations for the speech data set on a continuous scale. The top two graphs indicate the occurrence of the duration ratios, the middle graph shows the nPVI contribution for each ratio, and the bottom two graphs indicate the actual nPVI contribution of the duration ratios. The total proportions and the nPVI contribution of the occurrence of SL, LS and equal durational combinations are indicated in the graphs.

Discussion

Two factors which shape the nPVI were distinguished in the current study: an underlying distribution and a tendency of sequencing note durations. The question being examined was whether one factor contributes more than the other to yield the different nPVIs. The E-nPVI, which assumes random ordering, showed the similar amount of differences as for the nPVIs, suggesting that the nPVI difference was already determined to significant degree by the different choice of the note durations. The same effect was observed both for music and speech data.

The E-nPVI also revealed some tendency in sequencing note durations. When compared to the probability of occurrence predicted from the underlying distributions, the preference for less contrasted durational sequences was observed in all the music data sets, as E-nPVI was always higher than the obtained musical nPVIs. Similarly, less contrasted combinations occurred more often in the Japanese speech data, as E-nPVI was higher than the obtained nPVI for Japanese speech. However, more contrasted combinations occurred in the English speech data as E-nPVI was lower than the observation. It is hard to make a direct association between music and speech data regarding this preference in ordering durations. That being said, some commonalities between the domains were observed when more detailed characteristics in ordering durations, i.e. the occurrence of SL and LS duration combinations, were studied. The SL patterns occurred more than LS patterns in both English music and in English speech. The opposite was true in Japanese music: a significantly a larger number of LS patterns was observed as compared to SL patterns, with the occurrence of SL patterns being less pronounced in Japanese speech as compared to English speech. In the literature, a difference in long-short cyclic non-speech temporal patterns by North American and Japanese adults has also been reported (Iversen, Patel, & Ohgushi, 2004). In their perceptual study, US adults tended to group the temporal events into short-long patterns, while Japanese tended to group them into the opposite combination, long-short pattern. Interestingly, the pronounced occurrence of long-short intervals found in music and speech data in English in this study coincide with this finding. Though speculative, it is tempting to conclude that this interesting link between the perceptual result and the environmental statistics reflects how one learns the processing of temporal sequences through environmental exposure.

Another question investigated in the current paper is whether the musical nPVI method can include music from a culture that uses another type of language than syllable- and stress-timed. The musical nPVIs of the E set were expected to be higher compared to those of the J set, according to the tendency of their speech vocalic nPVIs. However, there was no significant difference when the E and J sets were merely compared, indicating that they share a similar contrast of successive durations in the rhythm of the melody. This result is rather surprising; music with lyrics is more likely to reflect the rhythmic characteristics of speech than instrumental music. For example, Palmer and Kelly (1992) showed the accent in prosodic structure of lyrics and music is tend to be aligned. Indeed, detailed analyses for the subset of the data in the current study indicated that language of the lyrics does have an effect on the musical rhythm. The musical nPVI for music by Japanese composers was significantly higher when they composed musical rhythms with English lyrics, as compared to the rhythms with Japanese lyrics. It was even higher than that of US music sets. This might be a similar phenomenon to what Patel and Daniele (2003a) showed in their study, that the nPVI value for one French composer (syllable-timed), whose style of composition aspired to the German music tradition (stress-timed), was extremely high. In this manner, the Japanese composer's admiration of English popular music might be reflected in a high nPVI value, as they tried to replicate what they consider the style of English popular music.

However, the language of lyrics is clearly not the only potential factor that determines the musical rhythmic organization because not all music contains lyrics; there is a possibility that the language of the mother tongue, which is not necessarily the language of the lyrics, is influencing the construction of rhythms. For example, in the previous studies, correlations have been shown between the rhythms between mother tongue and the instrumental music (Huron & Ollen, 2003; Patel & Daniele, 2003a, 2003b). The detailed analyses revealed that musical rhythms in the J_E set still showed some signs of a Japanese way of composing rhythm, such as less usage of short-long combinations than English. One particularly interesting finding was the low refrain nPVI for the J_E set, despite of its high nPVI when whole melody was taken into account. Although the exact reason of this finding is still up for debate, it might be the case that, when one tries to compose an "attractive", characteristic moment, such as a hook or a catchy phrase, the rhythmic structure becomes more varied in refrains, whereby it possibly reflects the characteristics of the mother tongue more.

It has been shown that huge variations in prosodic structure exist within the same language. For example, within British English speeches, diverse prosodic characteristics have been identified (Grabe, 2004). Especially noteworthy is the difference in rhythmic structure between British English and Singaporean English which has been shown using nPVI (Ling et al., 2000). Such local prosodic diversity within English speech might play some role in organization of musical rhythm, as the nPVI for E_{UK} and E_{US} was not similar.

Nonetheless, the size of the differences found in these musical nPVI studies are usually rather small as compared to the differences between British English and Japanese in speech, which is approximately 20 nPVI points. Patel and Daniele (2003) explained that small differences such as this are a reflection of the larger within-culture variability of musical nPVI values than those of speech nPVIs. This different variability might be based on the nature of durations in music and speech. In musical scores, as we used here, note durations are discrete and are precisely defined. A single music note can be quite long, while the durations of the vocalic or intervocalic intervals are not defined in the text and the phonetic duration cannot be as long as a musical note.

In this characterization, musical scores would be equivalent to written text in language data, while musical performance would be equivalent to speech utterances. Thus, the use of scores instead of musical performances for the nPVI calculation does not constitute a perfect match between domains. Different from written sentences, though, musical scores clearly specify the time span of note durations, which represent how the musical rhythms are likely to be organized in reality. Thus, scores must still reflect the rhythmic characteristics of music to a high extent. One consideration for future research is that the musical nPVI needs to be applied to performance timing in music for a better match to its application in speech.

The methodological contribution of this paper lies in showing that there is more to the nPVI than meets the eye. First, underlying distributions need to be checked before the nPVI difference can be interpreted as a mere preference for higher or lower contrast in serial order. Secondly, there may be some effect of sequencing durations, such as asymmetry, that escape from the nPVI analysis but can still indicate cultural differences.

What does the correlation between measures of rhythm structure of music and speech imply? Does it suggest the existence of shared cognitive path between language and music information, or an effect of exposures to a certain type of music on composers' creation in rhythm? The answer to this question is not easy. The cognitive link between language and music has been studied at different levels. Neurological findings suggest both association and dissociation between the two domains according to the activity of functional structures (Besson et al., 1998; Besson & Schön, 2003; Maess, Koelsch, Gunter, & Friederici, 2001; Patel, 1998; Peretz et al., 2004a). Other behavioral studies have demonstrated that lyrics and melody in songs are integrated in memory when one recognizes songs (Serafine, Crowder, & Repp, 1984; Serafine, Davidson, Crowder, & Repp, 1986), pointing to a direct link between the domains. Studies using music nPVI reveal the parallel in prosodic structures between domains, which could possibly be a product of a shared path in the cognitive architecture. However, it still might well be the case that composers learn to format the musical rhythms through an experience of the temporal patterns in their cultural environment. In the current study, we have shown that the rhythm in songs carries the characteristics of language of the lyrics. An exposure to music that reflects the characteristics of the mother tongue of the composer's culture, or to non-musical rhythms from speech might play a role in the composer's decision making in organizing musical rhythms.

Chapter 4

The Bayesian way to relate rhythm perception and production

Abstract

Measurements of the perception and production of simple rhythmic patterns have been shown not to be in line in some cases. In this study it is demonstrated that a Bayesian approach provides a new way of understanding this difference, by formalizing the perceptual competition between mental representations and assuming possible nonuniform a priori probabilities of the rhythmic categories. Thus we can relate the two kinds of information and predict perception data from production data. In this approach, the contrast between rhythm perception and production data, taken from different studies in the literature, was shown almost to disappear, assembling independent prior probabilities from counts of patterns in corpora of musical scores, or from a theoretical measure of rhythmic complexity. The success of this Bayesian formalization may be interpreted as an optimal adaptation of our perceptual system to the environment to in which the produced rhythms occur.

Temporal Patterns

Processing sequences of short time intervals plays an important role in our everyday life, for instance, in picking up stress patterns in conversation and in experiencing music. The importance of studying time relations as a mental phenomenon had already been brought up by the end of the nineteenth century (Jastrow, 1890). Since then, perception of time and action in time have attracted much empirical work (e.g., Fraisse, 1984).

Sequences of time points, marked by events, that is, clicks or onsets of notes, are the domain of these studies, though they are usually specified as a sequence of time intervals between events (inter-onset intervals). In musical scores, the notation of time intervals that constitute a rhythm is based on simple integer relations, and rhythm can indeed be represented as a sequence of integers. The term rhythm in this study will be used to mean such symbolic sequences. However, deviations from these perfect ratios in the performance of a musical score are usually large and cannot be interpreted entirely as noise. They partially constitute intended timing patterns that can communicate the structure of the piece (Sloboda, 1985). In this study, the term performance will mean a sequence of real-time intervals that carries both the rhythm and the expressive deviations.

Humans have a highly developed cognitive system for processing these sequences. The complexity of the mechanism stems from the fact that the two domains of information interact: a symbolic representation for coding rhythmic structure and a way to represent the small continuous deviations that make up the expressive performance. Note that the same rhythmic sequence can be played with various kinds of expression; for example, it can be made to sound swinging or laid-back by introducing small deviations from strict mechanical timing. Thus a notion of best, perfect, or ideal performance of a rhythm can never exist: it depends on the chosen style and the interpretation. Although a symbolic discrete code and continuous information are communicated when a rhythm is performed and subsequently perceived, both types of information become indistinguishable by being combined before they are transmitted as a sequence of time points through the same one-dimensional channel. Thus a large deviation in timing may very well upset the perception of the rhythmic structure itself.

There has been some evidence for categorical perception of rhythm. This process of perceiving the rhythmic structure for a performance is characterized by an increased sensitivity for detecting performance differences near the boundaries of the categories. Clarke (1987) conducted these experiments and showed the existence of categorical boundaries between specific rhythmic patterns. He also demonstrated that metric context (triple vs. duple) causes a shift in the position of the boundary. Schulze (1989) examined rhythmic categorization by using a different experimental setup, including varied tempi. He observed that subjects were able to identify the rhythms reasonably well, in spite of tempo variation. Desain and Honing (2003) specified the systematic mapping of continuous time intervals to rhythms for three-interval patterns and showed that the way categories are formed is affected by metric context. But in all these studies it is quite obvious that, while one perceives the rhythmic structure of a performance, the continuous information is still accessible, allowing one to perceive the expressive character of the performance.

In studies of music performance and expressive timing, it has been shown that there is no neutral, inexpressive way in which only the symbolic structural part of a rhythm can be communicated. Besides, expressive timing is not a random deviation from mechanical performance but has a certain regularity. In general, systematic deviations are usually observed (e.g., Gabrielsson, 1999, 2003) to be linked to the structural units in the piece (bars and beats, phrases, voices) (Clarke, 1985; Palmer, 1997; Sloboda, 1985). Several studies showed that playing an impassive performance, without any expressive deviation, is not even possible (Palmer, 1989). Repp (Repp, 1992; 1995; 1999c) has shown that deviations from a mechanical performance in accordance with expected regularities are harder to detect. These findings suggest that expressive timing is obligatory, inherent in the musical performance in a systematic way, and that our cognitive system seems to require it.

The Relation Between Rhythm Perception and Production

There have been many production studies in which rhythm has been characterized as expressive renditions of sequences of integers, either with strictly controlled experimental material (Gabrielsson, 1974), for full music performances (Timmers, 2002), or somewhere in between (Repp, Windsor, & Desain, 2002). The perceptual topic of the distribution of performance timing that allows for perception of a specific rhythmic structure has received less attention but has been investigated as well (Clarke, 1987; Desain & Honing, 2003; Schulze, 1989). Nonzero mean-time deviations from strict mechanical timing are commonly reported. It is surprising, however, that the reported means of the deviations from strict mechanical timing are often not consistent between perception and production studies. Consistency of perception and production would be an obvious assumption if we communicated with others and listened to ourselves while producing rhythm. Much classical work on the processing of rhythm perception and production has been based on this assumption (e.g., Eisler, 1976), but studies often focus on only one of the two processes, which might explain why the inconsistency has been long overlooked. In the past 20 years, however, there have been more studies in which the two processes were studied in conjunction (Drake, 1993b; Povel, 1981; Repp, 1992, 1995, 1998; Sternberg, Knoll, & Zukofsky, 1982), and they report that observed values of rhythm perception and production are not always consistent. For example, Sternberg et al. (1982) found that the durational ratios of perceived two-interval rhythms (using a perceptual judgment task) and those of produced rhythms are different, especially for short intervals (an example of discrepancy is presented in Figure 4.6). The perceptual deviations were found toward enhancing the contrast between two intervals while the production tendency was toward assimilation of two intervals. Thus, a perceived rhythmic category seems to occupy a separate region of the space of all possible performances as its performed counterpart. Taken at face value, this curious fact constitutes counterevidence for theories that postulate perception and production processes as closely integrated.

Sternberg et al. (1982) proposed a model in which rhythm perception and production tasks share a common analog representation but contain several internal transformations of the temporal patterns. The model does not require that the two tasks share these transformations, which accounts for the discrepancy. There have been, however, more claims associating characteristics of perception and production in understanding the cause of this discrepancy. For instance, some authors postulate, as reason for a deviation from the mechanical timing in rhythm production, that we compensate for peculiarities of our perceptual system; we might compensate for a perceptual tendency to hear intervals short by playing them longer (Drake, 1993b; Ihre, 1992; Penel &

Drake, 1998, 1999). Others claim the reverse: perception is constrained by production. For example, the learning of musical production evokes musical expectation that interacts with the way the listeners perceive temporal patterns (Repp, 1992). Yet another explanation states that perception and production are not in a causal relation, but each interact with the other in relation to the musical structure; both tendencies found in the perception and production are restricted by the musical structure itself (Repp, 1995). Yet none of these theories can adequately explain and predict the differences found so far.

A more fundamental issue may need addressing when rhythm perception and production are compared; it lies in the presence or absence of competition between the mental codes for varied rhythms and the possible nonuniform nature of the competition.

Rhythm Perception and Production Tasks

Rhythm perception and production are quite dissimilar tasks. In rhythm production, one mental representation of a rhythmic code is active. Its repeated realization, via a motor program, yields a distribution around a certain timing pattern. In rhythm perception, the space of possible timing patterns is probed. The stimulus is presented and a rhythmic code must be chosen in response. Several codes may be possible candidates for a certain stimulus. Thus, in a perceptual task the mental representations are in competition, whereas in a production task, the choice of the code to be activated is clear, as it is usually presented in the instruction.

Moreover, in production, rhythms are activated in a simpler way because only the target rhythm is selected and performed. In perception, however, mental codes for rhythms are in competition for selection as a perceived rhythm and as a more stable or simple representation, even if it constitutes a less close fit, or may prevent the choice of a closer but less stable fit. Furthermore, this competition may also be biased on the response side because selecting an unlikely rhythm, one not often heard, may be not an optimal choice. Therefore, certain, commonly occurring rhythms attract more responses (the areas in performance space that represent these rhythmic categories are larger) than others: the competition in perception may well be biased.

The difference in task characteristics can, in a fundamental way, influence the distributions of the empirical data. Consequently, comparing means and variances of perception and production data, as in Sternberg et al. (1982), may not lead to valid conclusions. We will introduce the necessary probabilistic method, Bayesian modeling, to address this issue and check this solution's performance on the empirical data.

Bayesian Modeling

A Bayesian Approach

In the Bayesian approach, the probability of a hypothesis being true, given an actual observation, is derived from the probability of the observation, given the hypothesis is true. In the calculation, the a priori probability of the hypothesis, in the context of all possible hypotheses, is taken into account. A Bayesian approach in perception and cognition was first introduced in signal detection theory, which was developed to investigate optimal strategies for the detection of signals in the presence of noise (Green & Swets, 1966; Tanner & Swets, 1954). Since the late 1970s, the quantitative application of the Bayesian approach has been applied in diverse areas of research.

The hypothesis that biological perceptual systems can be explained by using a Bayesian approach has been tested in the field of visual perception with much success (Knill & Richards, 1996). For example, the Bayes rule was used to give precise predictions about the perception of visual movement (Weiss, Simoncelli, & Adelson, 2002), and it provided a basis for the explanation of visual illusions (Geisler & Kersten, 2002).

The power of Bayes' rule has been fully exploited in Bayesian inference in more complex domains (these can be formalized by so-called graphical modeling; see Jensen, 2001). It has even been proposed as a general processing method for cognition, modeling upward and downward streams of information (Dayan, Hinton, Neal, & Zemel, 1995). Often an optimal (perceptual) strategy can be deduced. In our proposal, however, only a simple application of Bayes' rule is necessary to relate two conditional probabilities.

In producing a temporal pattern, a rhythm is provided, as symbolic code or musical score, and the conditional probability that a specific performance pattern arises, given this score, is estimated from repeated trials or from responses of a pool of subjects. In perceiving a rhythm, a temporal pattern is presented as performance, and the subject is required to identify the rhythm (the score). The conditional probability that a score is perceived, given this performance, is estimated from the responses. Bayes' rule relates these two quantities, formalizing the notion of nonuniform competition. It does so by the multiplication of the production distributions by, possibly nonuniform, a priori likelihoods for the rhythms themselves, followed by a subsequent renormalization. This transformation of production data should, according to Bayes' rule, be equal to the perception data, to be shown in detail later. Thus, the Bayesian relationship highlights both how rhythm perception and production data are the same—as one is derivable from the other—and how they are different, as the observed distributions are transformed versions of each other.

By considering the priors in a purely probabilistic interpretation, the familiarity of rhythms can be estimated by measuring frequency-of-occurrence information, for example, from corpora of musical scores and estimates of the amount of exposure of the subject to these music compositions. We will describe this task in detail later.

Taken in a pragmatic, nonprobabilistic way, priors may be used to reflect something else: some patterns are cognitively simpler or easier to code and memorize than others. Simplicity measure may not be similar to a familiarity estimate, because likelihood of rhythms may be expected to be related to complexity by a bell shape, as composers often avoid both too simple and too complex rhythms, as in visual art, where patterns with a medium complexity are appreciated as more interesting or beautiful (Berlyne, 1971; Birkhoff, 1933; but see Boselie & Leeuwenberg, 1985).

Although it is clear from a strict probabilistic stance that the concept of likelihood is necessary for a correct application of Bayes' rule, there remains the question of whether likelihood or simplicity is the most important concept for encoding mental representations (Helm, 2000). If simplicity is indeed the central factor in choosing among competing representations, what kind of structure can we expect the set of priors to have, and which temporal patterns can be considered simpler? We will review some of the literature on these issues.

Rhythmic Complexity in Perception and Production

It is well documented that temporal patterns that can be represented as small integer ratios are easier to process than ones needing higher ratios. Spontaneous rhythmic patterns, those produced without an indication of specific rhythmic structure and tempo, are typically made up of only two interval durations whose subsequent ratio is roughly 2:1 (Fraisse, 1946, 1956; see Clarke, 1999, & Fraisse 1982 for a summary). Also Povel (1981) has shown that the reproduction of two interval patterns (ranging from 1:4 to 4:5) was strongly distorted in the direction of 1:2. The same effect was found in other experiments (Essens, 1986; Essens & Povel, 1985; Summers, Bell, & Burns, 1989; Summers, Hawkins, & Mayers, 1986). Some studies show a preference for duple subdivisions over triple subdivisions (Drake, 1993c) predicting, for example, a higher simplicity for 1:2 than for 1:3. Furthermore, the difficulty in maintaining a clear distinction of duration patterns has been shown even when expert musicians are forced to produce a complex rhythm at a quick tempo (Peper, Beek, & Vanwieringen, 1995; Repp et al., 2002). This result can be predicted by a theoretical account of the complexity of ratios.

One problem in formalizing a notion of rhythmic complexity is the interaction between the rhythmic structure of the pattern (intervallic structure) and its metrical interpretation (hierarchical structure), an aspect often implicit in, or induced by, a temporal pattern. Timing of the production in musical performance usually varied depending on the position in the metric context (e.g., Gabrielsson, Bengtsson, & Gabrielsson, 1983). Many approaches to rhythmic complexity combine information of theoretic and perceptual factors (Pressing, n.d.; Shmulevich & Povel, 2000; Tanguiane, 1993). Alternatively, derived or indirect measures could be considered, including the amount of syncopation. For instance, Longuet-Higgins & Lee's (1984) measure of syncopation strength indicates the amount of syncopation of a rhythmic pattern, given a certain metrical interpretation. A more syncopated pattern could be considered more complex. Nevertheless, because these theories define complexity within a given meter, they cannot be used in our study because the data in the experiments were obtained without control of meter.

There are other hypotheses regarding the complexity of ratios of temporal patterns, which may be formalized as a set of priors in a Bayesian approach. The purely numerical notion of a so-called Farey tree, which is sometimes applied to explain the human ability to process different temporal ratios, is a good candidate. For example, Peper et al. (1995) demonstrated transitions of the ratio of different tapping rates realized by both hands at the same time ranking of the complexity of ratios, according to the depth in a tree. The complexities increase as we move from the root (see Figure 4.1). Here we present hierarchical ratios defined as the duration of first interval divided by the total duration of the pattern (e.g., 1/2 signifies two equal durations, i.e., 1:1, and 3/4 is used instead of 3:1). Note that in this manner the hierarchical durational ratio is always between 0 and 1.

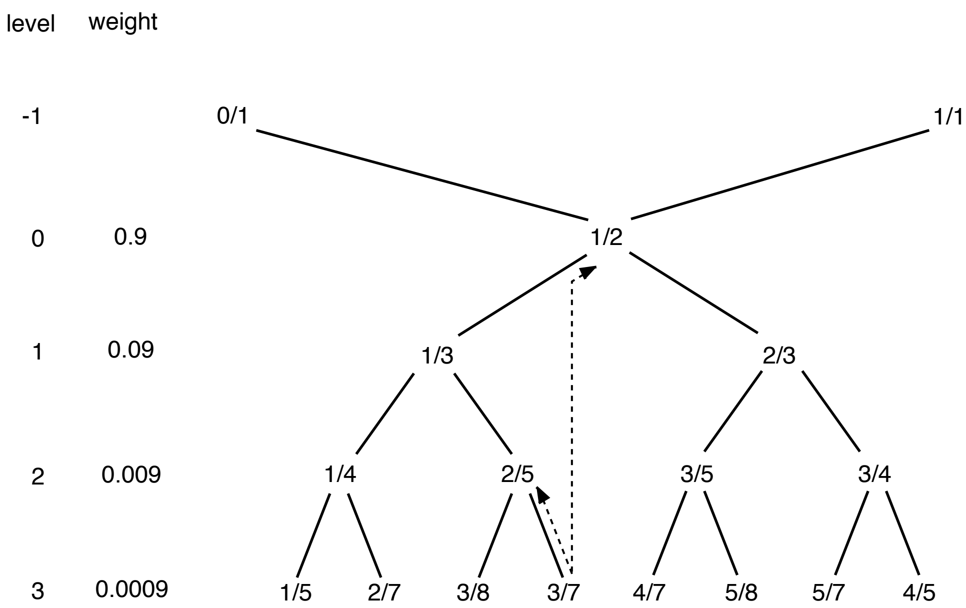


Figure 4.1 The Farey theory of the hierarchical ordering of the ratios according to their complexity, visualized as a tree structure. The Farey tree provides a structure of rational numbers, which can be derived algorithmically (Cvitanović, Shraiman, & Söderberg, 1985; González & Piro, 1985). The ratios at each level (m"/n") in the tree are obtained from two parent ratios located at a higher level of the tree (m/n and m'/n'), m"/n"=(m+m')/(n+n'). One parent ratio is connected directly to the daughter ratio by a branch (see example arrow). The other parent ratio is found following the vertical arrow upward till it crosses a branch, and then following the branch upward.

Next to the theoretical notion of the Farey tree, we need to look at familiarity of rhythmic patterns. Although estimating a subject's prior exposure to various rhythms is an impossible task, counting the rhythms in a corpus of musical scores may be taken as a first approximation to likelihood of a rhythm. As the rhythm perception and production data are usually gathered for a fixed number of notes within a repeating time interval (beat or bar), the counting in scores has to account for only n-note patterns that span the unit. Also, the indicated tempo in the score guides the selection of the metric unit under consideration, as it should be roughly the same duration as the unit used in the perception and production experiments, for tempo matters in both rhythm perception and production.

Two interval rhythms, notation and formalism

For clarity the formalisms used will be based on two interval temporal patterns. Generalization to higher dimensions is straightforward. Let us first characterize a two-interval score rhythm. Assume three successive rhythmic events (note onsets) at score-time z_1, z_2 and z_3, counted in arbitrary units ($z \in N$). These three points specify two successive note durations ($z_2 - z_1$ and $z_3 - z_2$) and one hierarchical ratio $c = (z_2 - z_1)/(z_3 - z_2)$ of the first note's duration interval regarding the duration of the whole sequence. Each possible rhythm in this domain is thus uniquely identified by a positive rational ratio $c \in Q$ with $0 < c < 1$, the rhythmic code or category, and we will use this ratio c as the name of a rhythmic structure, irrespective of the notational level (e.g., both the sequence of two quarter notes and of two eighth notes form the ratio 1/2).

Next define a two-interval performance t. Assume three successive temporal events (onsets) at real-time x_1, x_2, and x_3 (e.g., in seconds). These three time points specify two successive inter-onset intervals ($x_2 - x_1$ and $x_3 - x_2$) and one hierarchical ratio $t = (x_2 - x_1)/(x_3 - x_1)$ of the first interval with regard to the duration of the whole sequence. Each possible performance is thus uniquely identified by a real ratio $t \in R$ with $0 < t < 1$. We will use this ratio as a label for a performance.

In a production task, a rhythmic structure c is provided as stimulus or instruction, and a performance t is produced as response. In a perception task, a performance ratio t is presented as stimulus, and a rhythmic ratio c is required as response.

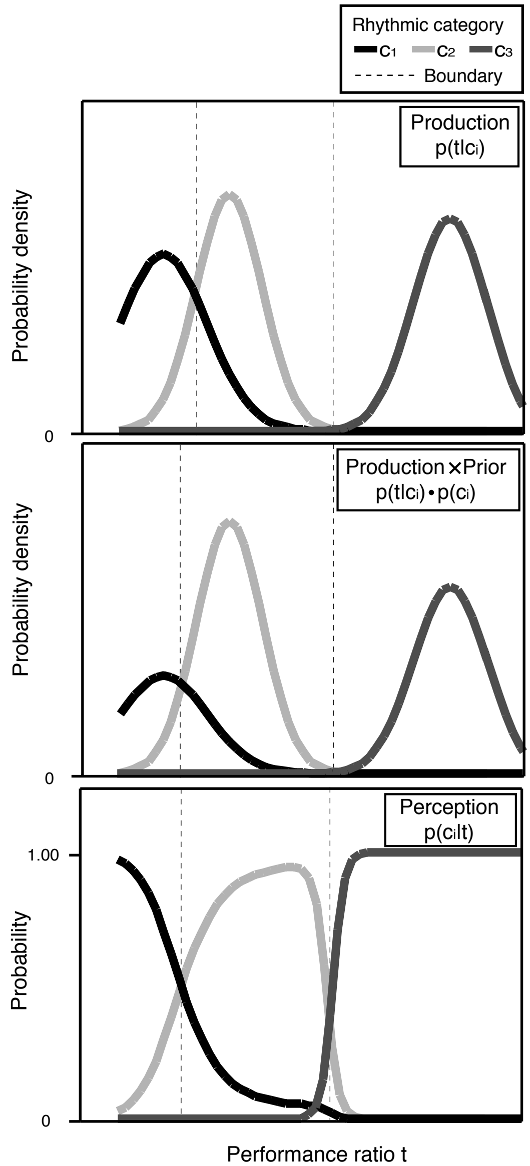


Figure 4.2 **a.** Example distributions of production data: p(t|c). **b.** Example distributions of production data multiplied by priors: p(c|t) × p(c). **c.** Example distributions of perception data: p(c|t) (See formula 1).

A production data set consists of a number of probability densities over the domain of performance ratios, one for every rhythm c considered, which is illustrated in a schematic way (top panel of Figure 4.2). Each curve represents the probability for a specific performance t given the instructed rhythm c. Thus, this data set specifies $p(t|c)$[1], the conditional probability of a performance given a rhythmic instruction. Note that as the curves are densities, the surface area under each equals 1. Because the raw data of a production experiment consist of sets of t collected for each c, the density curves must be estimated from these sets by constructing a histogram or fitting a theoretical continuous distribution.

A perception data set consists of a number of probability curves over t h e domain of performance ratios, one for every rhythm c (bottom panel of Figure 4.2). Each curve represents the probability for a specific response rhythm c, given the presented performance t. These curves are like the receptive fields in visual perception theories, or the tuning curves in the domain of auditory perception. Thus, this data set specifies $p(c|t)$, the conditional probability for a perceived rhythm c given a performance t. Note that as the curves are probabilities, not densities, here their sum equals 1 for each value of t. Thus, at each t the perception data specify a discrete probability density over the responses. Because the raw data of a perception experiment consist of sets of c collected for each t, the probabilities simply reflect (are estimated by) the response proportions.

A prior data set consists of a set of a priori likelihoods of occurrences, one for each rhythmic ratio c. It is notated as $p(c)$ and reflects the possibly nonuniform exposure to various rhythms.

With these definitions in place, and using Bayes' rule, it is possible to define the relation between the constructs. Bayesian modeling provides a framework for reasoning with uncertainty. The notion of conditional probability is central and denoted as $p(a|b)$, which expresses the probability of a occurring when it is given that b occurs. Bayes' rule relates the probabilities $p(a|b)$, $p(b|a)$, $p(a)$, and $p(b)$. Applying it directly to our case, the rule dictates:

[1] It would be more formally correct to p resent t h e conditional probabilities as $p(T| C=cj)$, thus a Bayes rule as $P(C=cj| T) = p(t| C=cj) \times p(C=cj)/p(t)$. However, we opt for the shorter notation.

$$p(c|t) = \frac{p(t|c) \times p(c)}{p(t)} \qquad (1)$$

This can be read as: *the probability of a rhythm being perceived, given a (presented) performance, is equal to the probability of that performance being produced, given that rhythm (as instruction), times the prior probability of that rhythm, divided by the probability of the performance arising in any case*[2]. The latter term sums over all possible cases (any rhythm). It acts as normalization constant and can be rewritten as

$$p(t) = \sum_i p(t|c_i) \times p(c_i) \qquad (2)$$

To return to Figure 4.2 for an illustration of this calculation, each production density curve $p(t|c_i)$ from the top panel is scaled by a prior probability $p(c_i)$. This yields the middle panel of Figure 4.2. Then the curves are renormalized, making them sum to 1 for each value of t by dividing by their sum. This maps these likelihoods to the proportions of (forced) responses in the bottom panel, which is taken to predict the perceptual data.

Surprisingly simple, Bayes' rule may thus be able to give an explanation for the differences occurring in the means and variances reported for perception and production, as it explains the transformation of the shape of these curves. Looking at Figure 4.2, we can see that for each curve (i.e., each rhythm) the performance mean and variance in the two data sets differ, because for perception a strong competing neighbor on one side may skew the response curve.

Before we embark on testing, if the formalism can work on real data sets, there is one caveat. This method can be used only for performance ratios where production data exist ($p(t) > 0$). No prediction can be made for the perception of a performance for which the probability produced is zero

[2] In this formulation, considering the environment, to-be-perceived and produced performances have been equated. Considering the mental representation of rhythmic structures, task instruction (production) and task responses (perception) are equated as well.

for all instructed rhythms. Thus, many rhythms need to be considered in the production experiment. Furthermore, this set should contain all rhythms obtained as responses in the perception task; or vice versa, the possible responses in the perceptual task should be limited to the set of rhythms tested in production. By adding the necessity of an equal tempo in both tasks, these limitations created the difficulty of finding the appropriate data sets for this metastudy.

Hypotheses

Bayesian inference can provide a new way to interpret data, by stating that perception and production are only apparently different, because the difference results from the sensitivity of the rhythmic categories to (nonuniform) competition in perception. Stated in other terms, we hypothesized that perception data predicted from production data using Bayes' rule are closer to observed perception data than the production data as a whole.

We will first elaborate our hypotheses in these terms before introducing a more rigorous test. Because of the disparate nature of perception and production data, the statistical test of difference can be carried out using only a rough indication of similarity such as correlation. This goodness-of-fit measure can indicate the closeness of production data to prediction perception data that use various priors. We use a general two-dimensional (rhythm * performance) correlation measure, which is computed from corresponding two variables over all categories and all time points. It gives us the amount of variance in the perception data explained by the production data as well as predicted perception data.

The direct comparison gives us the amount of variance in the perception data directly explained by the production data (r_d^2). Since there are different sets of priors, the prediction using Bayes' rule comes in several variants. A first variant poses uniform priors in which all ratios are treated equal. The fit between the perception and this uniform prediction (r_u^2) can be interpreted as an indication of the success of taking into account only competition. The second option is a nonprobabilistic interpretation using a complexity measure, the Farey tree, giving us r_f^2. In the next variant, the

priors are derived independently from frequency counts in three different corpora of musical scores yielding r_{sA}^2, r_{sE}^2 and r_{sT}^2 (Anthem, Essen, and Theme, respectively). First, we expect r_d^2 to be poor when compared with other Bayes predictions. Second, since a uniform prior does not differentiate between rhythmic categories, we assume uniform priors cannot be as good as score priors and Farey priors: $r_u^2 < r_s^2$ and $r_u^2 < r_f^2$. The relation between estimated perception data by score count priors (r_s^2) and Farey priors (r_f^2) is unsure, as the reliability of the estimation of exposure from a corpus of musical scores is not known and neither is the perceptual plausibility of the simple numerical complexity rule. For the final variant, the priors are treated as parameters whose value is found by optimizing the fit between predictions and observations, yielding r_o^2. This option introduces many parameters, one less than the size of the set of rhythms, and it is obvious that this will result in the best fit: $r_s^2 < r_o^2$ and $r_s^2 < r_o^2$.
For a rigorous test of the significance of the difference between predicted and observed perception data, we applied the Kolmogorov-Smirnov goodness-of-fit test for each performance ratio t. The test examines whether the proportion of probability curves t (4/19 sec, 5/19 sec, etc.) between predicted perception data and observed perception data is significantly different[3]. Our hypothesis is that better predictions yield fewer points n at which the predicted probability of responses is still distinguishable from the observed proportion. The raw production data using this test cannot be related to perception. Thus, for a given significance level we predict $n_s > n_o$ and $n_f > n_o$. The r_o and n_o will be used as an estimate of a ceiling of the method's success: the maximally achievable congruence between a perception and production data set, using only the Bayes rule.

[3] The more common chi-square test cannot be used in this study because there were always categories for which the probability is zero. The underlying variable for the Kolmogorov-Smirnov test is basically required to be continuous, but the known violation of this assumption leads to only slight errors on the conservative side (Hayes & Winkler, 1970).

Table 4.1 Characteristics of the data from the five experiments compared in this study.

Mode and Data set	N	R	1/8 1:7 125	1/7 1:6 143	1/6 1:5 167	1/5 1:4 200	1/4 1:3 250	1/3 1:2 333	2/5 2:3 400	1/2 1:1 500	3/5 3:2 600	2/3 2:1 667	3/4 3:1 750	4/5 4:1 800	5/6 5:1 833	7/8 7:1 875
			Hierarchical and successive interval ratios, and first interval (ms)													
Perception																
[Desain & Honing]	17	1	166.9 (19.6)	-	169.8 (21.9)	197.7 (22.8)	270 (67.4)	338.6 (60.8)	421	488.5 (44.1)	579	636.6 (47.1)	731.2 (47.4)	-	812.5 (26.2)	818.9 (32.1)
[Sternberg (J2)]	3	3-5	59.3 (8.3)	79.7 (10.4)	105.4 (20.4)	154.4 (30.9)	207.3 (50.1)	303.6 (50.8)	-	451.7 (60.7)	-	-	-	-	-	-
Production																
[Repp]	12	6	-	-	-	-	278.6 (27.8)	315.6** (33.6)	364.1* (33.9)	499.6 (17.6)	508.6* (43.5)	631.0** (37.2)	709 (34.3)	-	-	-
[Sadakata]	12	8	-	-	174.3 (23.1)	230.6 (26.8)	267.1 (16.0)	332.5 (12.5)	-	500.9 (8.3)	-	646 (12.9)	721.8 (14.9)	753.2 (21.2)	809.1 (20.8)	-
[Sternberg (P4)]	3	250-1000	156.8 (30.4)	181.4 (30.2)	190.4 (30.0)	-	256.7 (20.3)	-	-	500.1 (20.6)	-	-	743.9 (30.2)	-	814.2 (30.1)	853.7 (40.8)

Note. N = number of subjects, R = number of repetitions. The mean interval duration for the first interval of each ratio is given in ms, and their standard deviations are shown in parentheses. An asterisk (*) indicates an interpolation method was needed; two asterisks (**) indicate the use of an extrapolation method to arrive at the appropriate tempo.

Application of the Method

Material

To yield a relevant comparison across situations with a different experimental method, a careful selection of the data sets was needed. The data sets used were collected from the study by Repp et al. (2002, production), Sadakata, Ohgushi, and Desain (2004, production), Desain and Honing (2003, perception), and Sternberg et al. (1982, perception and production), respectively. The task is described in Figure 4.3. Detailed data descriptions appear in Appendix 1.

All of the data sets used rhythmic patterns consisting of two intervals whose total duration was 1 second. In the perceptual studies, the subject is presented with a possibly repeated auditory pattern, and the task is to identify a rhythm. In the production studies, subjects were asked to perform a rhythm as a movement pattern by hitting a drum or by playing a piano. (See Table 4.1 for a list of the rhythmic ratios available in the studies). In Table 4.1, the means and standard deviations are also listed as calculated from the raw data (i.e., actual responses), or as taken from the original article (in case of the Sternberg et al., 1982, data set[4]). Though many studies show that the actual time durations (tempo) influence musical performance, the issue of time scale cannot be considered systematically in our study, as it proved impossible to obtain access to data sets with more than one tempo condition in common. Thus, although the individual studies may address other tempi, we restricted our analyses to one (moderate) tempo: patterns of two time intervals summing to 1 second. In a few cases, the data were not available at the exact tempo, and a small interpolation was needed.

As these data stem from quite separate experimental setups and the procedures and the musical character or naturalness of the tasks was quite diverse, we will first outline the tasks that the subjects had to perform.

In Repp et al., the pianists were involved in a natural musical undertaking: performing monophonic melodies on a piano at a given tempo. The Sadakata et al. experiment was somewhat artificial, by performing in a mechanical way a repeated drum pattern on a pad. For the perception experiments, the free transcription task of Desain and Honing was close to

[4] Sternberg, using moments, translates the "between" measurements into category means and standard deviations (see Appendix C in the original article).

everyday musical activities of musicians and composers. Skilled musicians participated in both the perception and production tasks of Sternberg et al. They identified the rhythmic category of the presented rhythmic pattern by specifying time duration in the perception task and tapped the rhythmic category along the metronome click in the production task.

A Priori Likelihoods: Empirical

To differentiate between rhythmic patterns, frequency counts were derived from databases of musical scores to serve as sets of priors. A very large corpus from a diverse kind of music is necessary to obtain appropriate counts of rhythmic pattern frequencies for the wide range that occur in music. For this, the frequencies of 14 ratios, which occur within metrical subdivisions, were counted from three diverse kinds of music corpora named Anthem, Essen, and Theme (a detailed description of each database appears in Appendix 2). Some ratios, such as 2/5 and 1/7, did not occur at all and reflect that divisions in 5 and 7 are much less common in Western music (London, 2001).

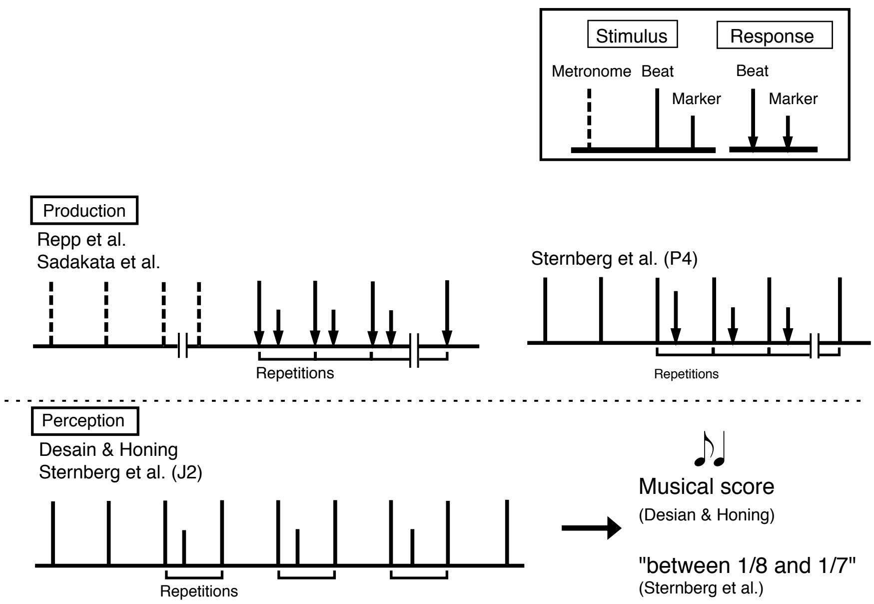


Figure 4.3 Example of the stimuli and produced responses in the experiments.

POCO (Honing, 1990) was used to collect the counts. As the empirical data deal with rhythmic subdivisions of a repeated unit of 1 second, which when presented or performed assume a metrical character, only note pairs that spanned a metric unit (bar, beat, or subbeat) were considered for the counts.

The frequency of occurrence of the ratios used in this study as they appeared in the databases is shown in Table 4.2. The total number of counted ratios was about 19,000 for the Theme data, 95,000 for Essen, and 4,000 for the Anthems. The range of the counts spans a large range: five orders of magnitude. The frequency of ratios not included in this study (shown as "other" in the table) is very small: .4% for Theme, .002% for Essen, and 0.005% for Anthem.

This result shows that almost all of the relations of two intervals can be classified into the 15 categories used. Table 4.2 brings out the amount of similarity between the very different corpora. The correlation between counts of two of the databases is always above .78 with a maximum .99.

Table 4.2 The frequency of all the ratios used in this study as they were extracted from corpora of musical scores: the Essen Folksong Collection [Essen], the Anthem set [Anthem] and Barlow & Morgenstern's The Dictionary of Musical Themes [Theme].

	Frequency		
Ratio name	Essen	Anthem	Theme
1/8	.000	.000	.001
1/7	.000	.000	.000
1/6	.000	.000	.000
1/5	.000	.000	.000
1/4	.001	.001	.012
1/3	.009	.002	.005
2/5	.000	.000	.000
1/2	.720	.473	.704
3/5	.000	.000	.000
2/3	.122	.012	.081
3/4	.140	.496	.175
4/5	.000	.000	.000
5/6	.006	.001	.008
7/8	.000	.015	.011
Other	.000	.000	.004

Note. .000 should be read as < .0005.

Complexity Measure as Prior

Though devoid of a probabilistic interpretation, any measure that assigns different weights to rhythms can be used as if it were a prior, as long as the measures are positive and sum to 1. In this way, we evaluate the Farey tree, a specific simple ranking of ratios according to their numeric complexity. As the tree (Figure 4.1) specifies only a ranking, and not a numerical value, we assigned the root (1/2) the maximum weight and assumed the weight at each level to be a fraction of the weights of the next higher level. We required the weights of the levels used to sum to 1[5]. This uniquely determines the weights.

Method

Elaborate steps were taken in the computations, partly because the data were not collected with the aim to compare them in the original studies.

Individual observations of *time intervals* are available for each subject, and each ratio was averaged over repeated trials for Repp et al. and Sadakata et al. A schematic of the procedure is presented in Figure 4.4. The main flow of the information is from left to right. The production observations were modeled by a beta distribution, which can describe these observations quite well, as they are range-limited and usually skewed (See Appendix 3 for more information about the beta fit). The fit was done using log-likelihood optimization. The discrete set of probabilities was calculated from the beta distributions using bins around the time grid of the perceptual data as the input for the Bayes calculation. The other input is a set of priors, for which some variants are available (uniform, three score counts, Farey tree). Bayes' rule outputs predicted perceptual judgment distributions, which can be compared with the observed perception data (see the bottom row of Figure 4.5). The comparison is made by calculating the correlation between distributions, as well as the Kolmogorov-Smirnov goodness-of-fit test. The resulting fit provides the evidence for conclusions about the hypotheses. Furthermore, the mean square error between distributions provides the measure to minimize searching priors that optimally predict the perception

[5] The ratios 0/1 and 1/1 were not taken into account, as they do not specify a two-interval pattern.

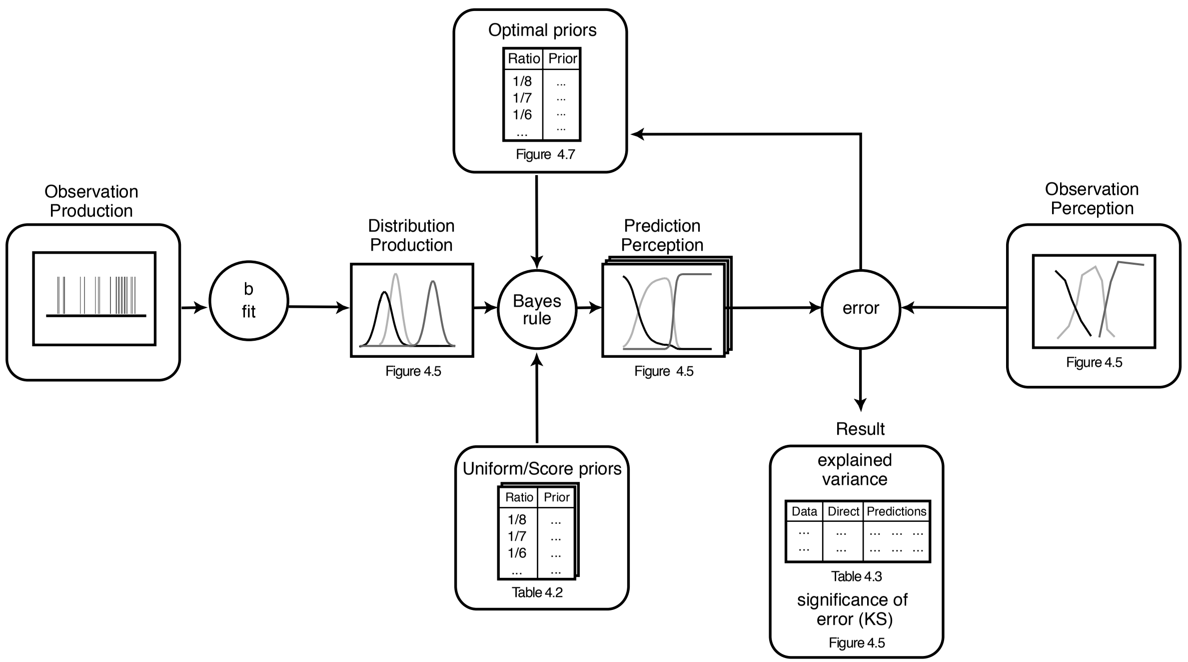


Figure 4.4 The paradigm used to compare perception data with production data using Bayesian modeling and the figures and tables in this article. The inputs are the raw production data (leftmost box), raw perception data (rightmost box), and a set of priors. After fitting a distribution to the production data and applying Bayes' rule, using one of the prior data sets, the perception data are predicted. The explained variance of the fit, along with the statistical significance of the remaining difference, is one of the results. The other result, a table of optional priors, is obtained when the fit is optimized using priors as parameters.

data from the production data. This optimization was constrained by requiring the priors to sum to 1.

The processing for the Sternberg et al. data set shares much of the information flow in Figure 4.4. Unlike Repp et al. and Sadakata et al., only the summary statistics, such as means and standard deviations, are available in this case. Thus, the production distributions had to be reconstructed using a symmetric beta distribution, approximating the given mean and standard deviation. Using the various sets of priors, we predict the perceptual data. However, the distribution of the target, that is, the observed perception data, cannot be reconstructed from the data in the study, as the relative proportions of responses for each ratio category are not available. Thus, we have to resort to deriving the predicted means and standard deviations and comparing them with the observed ones. Minimizing the difference (rms error) between predicted and observed means leads to a set of optimal priors. The results presented by Sternberg et al., using a direct comparison of perception and production statistics, and bypassing Bayes' rule, are considered the baseline.

In the case of the Desain and Honing perception study, a few outliers, that is, single responses isolated from the other responses for the same rhythmic category, were observed in the categories 1/6, 1/4, 1/3, and 5/6 (2%). They were treated as errors and excluded from the data. As rhythms used among studies do not completely agree (see Table 4.1), we selected the rhythms used in each production study for the corresponded perception data to be compared. As a result, a small amount of the Desain and Honing perception data had to be discarded and normalized in the comparison with Sadakata et al. study (3%) and Repp et al. study (6%) respectively, which result in the different shape of the Desain and Honing perception distributions in Figure 4.5. Thus, the perception data were normalized variously in each case according to the categories used.

From the Repp et al. production study, we could directly use the rhythmic patterns 1/2, 1/4, and 3/4 of the tempo condition normal (total duration is 1000 ms). Yet (linear) interpolation between the "Slow" and "Moderate" condition had to be used for 2/5 and 3/5 patterns, and extrapolation from the "Moderate" and "Slow" condition for 1/3 and 2/3 patterns. From the Sadakata et al. study, all intervals were available at the required tempo. Nevertheless, for both studies the performance tempo was not enforced and drifted slightly over repeated productions of the *time intervals*. As these drifts were very small at this moderate tempo, in the order of 3%, we normalized the *time intervals* to make them sum to exactly 1 second. The preprocessed data were entered into the next stage.

Results

[Repp] and [Sadakata] vs. [Desain & Honing]

Figure 4.5 shows how, after applying Bayes' rule, the distribution predicted from two production data sets approximates perception data. The results using production data from Repp et al. [Repp] are provided in the left column, and the results for Sadakata et al. [Sadakata] in the right column. In both cases, the original production data are shown in the top rows, and the perception data of Desain and Honing in the bottom rows. The second row gives predicted perception data obtained by applying a uniform prior; in the third through sixth rows, the results from various score priors and the Farey prior are presented. The vertical axis shows probability density at the top row and probability for predicted perception and real perception data. The horizontal axis shows the hierarchical ratio on a grid of 1/19, in accordance with the stimulus sampling used in the perception study. Note that a limited range is presented on the x-axis, as perceptual data are available only in that interval. The result of the Kolmogorov-Smirnov goodness-of-fit test is represented as a bar under each prediction. If the prediction on a certain performance ratio is significantly different from perception data, the bar under this time point is gray (p<.1, <.05) or black (p<.01). Nonsignificant difference, which indicates good predictions, is represented as white.

First notice the contrast between top and bottom rows: the nature of the tasks in rhythm production (top) and perception (bottom) is reflected in the divergent curves. It can be easily understood that conflicting means and variances are reported, given that the distributions themselves are so distinct. This dissimilarity is also reflected in the low amount of variance explained by production data (r_d^2), as given in Table 4.3 under "direct comparison". Now consider the second row: here are the predicted perceptual data using

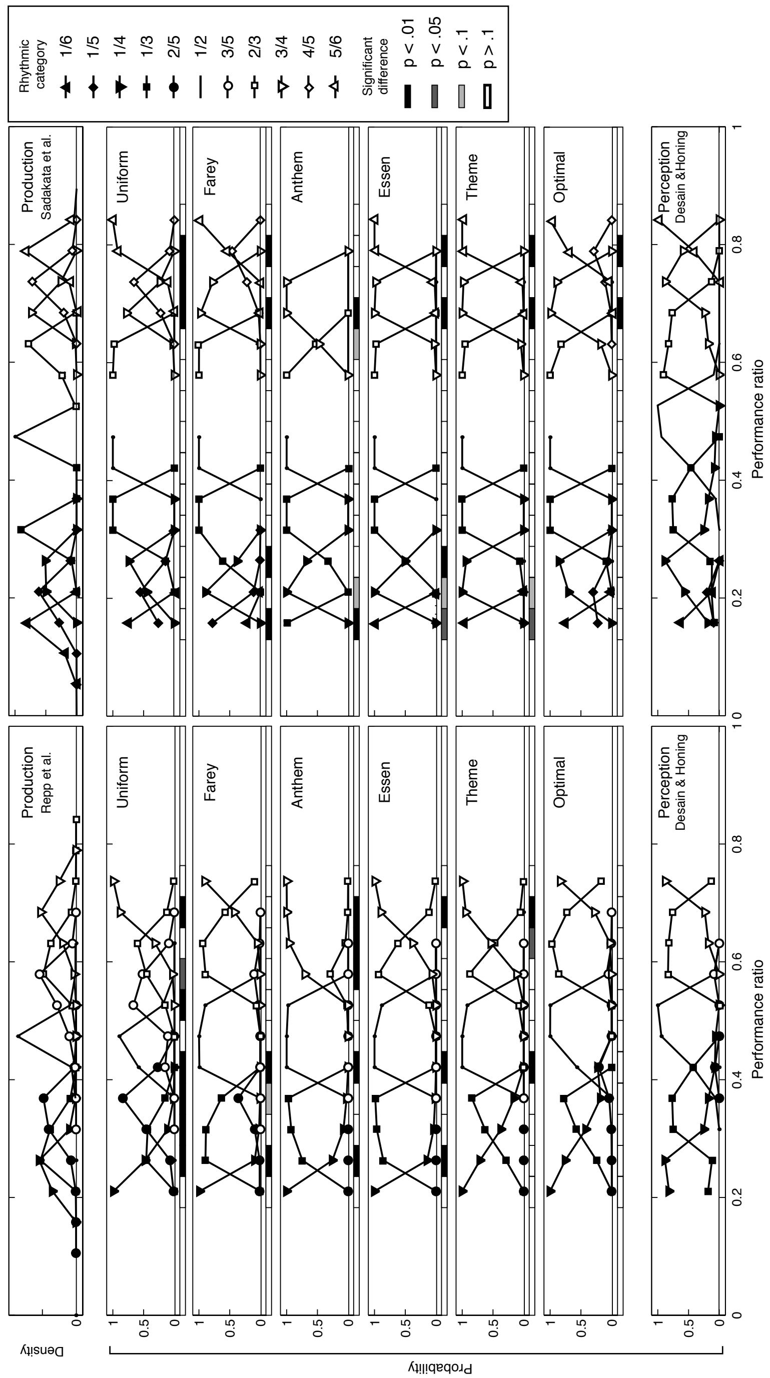


Figure 4.5 The observed and predicted distributions of rhythmic categories. On the horizontal axis the duration of the first interval is given; on the vertical axis the probability is represented, either of producing this interval given a rhythmic category or of judging this interval as a proper representation of the given rhythmic category. In the middle rows the perception data as predicted by Bayes rule with different priors are presented.

Bayes' rule with uniform priors, assuming all categories equally likely. Accounting for the different nature of the tasks regarding competition produces a considerable change of the shape of the distributions. The success of using uniform priors was varied between the data sets; the uniform priors explained the relation in the case of [Sadakata] quite well, whereas a considerable difference was still observed for [Repp]. The predicted perception data with Farey priors and with score count priors are shown in the next four rows of Figure 4.5. In the majority of cases, these priors provide fine predictions (see Table 4.3). The limits of the method are represented in the last row of Figure 4.5; the priors were considered as parameters and optimized for best fit, significantly raising the proportion of variance explained for [Repp] and [Sadakata].

Although the Farey tree seems counterintuitive in some respects (e.g., 1/4 is intuitively less complex than 2/5), priors from score counts and the Farey tree seem to reflect the relative importance of rhythmic categories to an extent, as they succeed in providing good r^2s in both production data sets. Yet, it has been shown that the fit can still be much improved at least in the [Repp] set, by optimizing the priors.

Kolmogorov-Smirnov goodness-of-fit tests (see Figure 4.5) showed that the number of points at which there is a significant difference between predicted and observed perception is considerably decreased by using the optimal priors in both data sets. Thus, the order on the appropriateness of the priors was as expected in [Repp], $r_d^2 < r_u^2 < (r_s^2, r_f^2) < r_o^2$ and $n_u > (n_s, n_f) > n_o$. But in [Sadakata], the order was dissimilar because uniform priors worked well, $r_d^2 < (r_u^2, r_s^2, r_f^2) < r_o^2$ and $(n_s, n_f) > n_u > n_o$.

Table 4.3 Proportion of the variance in the perception data explained by the production data (and Bayes' rule).

Data sets		Direct	Bayesian model with priors					
Production	Perception	r_d^2	Uniform r_u^2	Farey r_f^2	Score priors r_s^2 Anthem	Essen	Theme	Optimal r_o^2
[Repp]	[Desain & Honing]	0.34	0.38	0.71	0.47	0.65	0.73	0.93
[Sadakata]		0.39	0.66	0.62	0.50	0.68	0.72	0.74

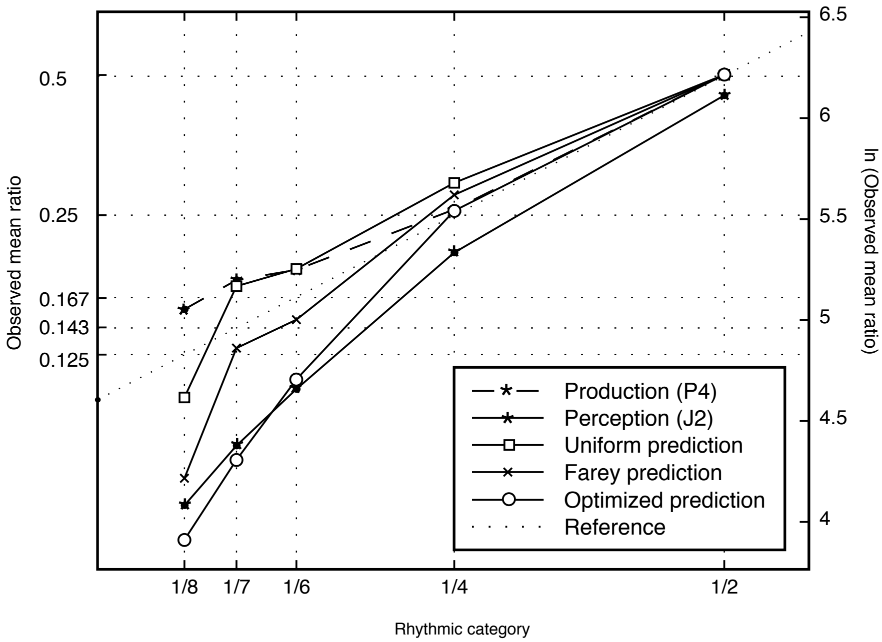


Figure 4.6 Observed mean ratios in Sternberg et al. production (P4) and perception (J2) data, and the mean ratios of perception as predicted from the production means by Bayes' rule using various priors on a logarithmic scale.

[Sternberg] vs. [Sternberg]

The row data are necessary for a thorough application of Bayes method; however, using approximations, we can still test if Bayes method works when only statistics (means, standard deviations) are known, for instance, in the study of Sternberg et al.

We explained in the method section that the data distribution of the production experiment (P4) for each ratio was approximated by symmetrical beta distributions. Means and standard deviations of the predicted perception data were calculated, using uniform priors, Farey tree priors, score count priors (Theme, Essen, and Anthem, respectively), and optimal priors. We compared these statistics with the judgment perception experiment (J2). Figure 4.6 shows the means in the same format as Figure 4.5 of the Sternberg et al. article. Note that Figure 4.6 was based on the average response of three participants from the original article, but the response by only one participant was plotted in Figure 5 in the original article. Observed and predicted mean values are plotted against the rhythmic ratio on a log scale. Exact timing provides the reference as the diagonal dotted line in this figure. As reported by Sternberg et al., there is a remarkable discrepancy between the results P4 and J2 (see data marked with * in Figure 4.6), constituting a contraction of the first interval for the perceptual task and an elongation of the first interval for the production task. This discrepancy becomes especially large for a ratio smaller than 1/4.

In Figure 4.6 one can see how the Bayesian approach derives predicted perception means from the production data, using different priors. Using the rms error (*e*) between predicted means and observed means (Sternberg, J2) as criterion, we found similar results as in that of [Repp] and [Sadakata] with $e_o < (e_s, e_f) < e_u < e_d$. The *e* were smallest when priors were optimized (e_o = .07), followed by Essen (e_{sE} = .10), Farey (e_f = .11), Theme (e_{sT} = .13), and Anthem (e_{sA} = .16), then Uniform (e_u = .17), and Direct comparison (e_d = .18). An excellent prediction was made by optimal priors, which makes the distinction between perception and production means almost disappear.

It is important to note that there are few rhythmic ratios contained in the Sternberg study; for example, there is no 1/3 between 1/4 and 1/2 and there is an absence of ratios smaller than 1/8. Note also that results of the score count priors are omitted in Figure 4.6, as not all predictions could be made due to the lack of these ratios in the score databases.

Priors

Comparing the various priors, reflecting the nonhomogeneity of the space of rhythmic structures, is one point of interest in our study. It is necessary for a good coverage of rhythmic categories to run a good optimization and derive a set of optimal priors. The score counts imply that rhythmic categories included in both the study by Repp et al. and the study by Sadakata et al. cover most of the musically reasonable ratios that can occur within 1000 ms (98.8% of Theme, 100% of Essen, and 98.5% of Anthem). The priors obtained by the combination of production data by [Repp] and [Sadakata] are of interest because they jointly cover these musically reasonable ratios. The data sets of these two studies were combined to arrive at a complete set of optimal priors [Repp-Sadakata].

In Figure 4.7, the theoretical Farey complexity of ratios, the proportion of occurrence as measured in the score counts, and the optimal priors obtained from [Repp-Sadakata] are presented on a logarithmic scale, to allow for the wide range. In the case of a zero prior, the corresponding value in the logarithmic graph was set to be the lowest rank. At first view their proportions appear to be quite distinct. As the distributions do not extend widely, the size of a prior determines the shape of the curves relative only to its direct neighbors. Thus the optimality of the priors m ay reflect only the local relation, and the global structure of the set may not be well expressed, that is, the relative proportion of the leftmost and the rightmost prior may be subject to a much larger estimation error than the relative proportion of two neighboring ones.

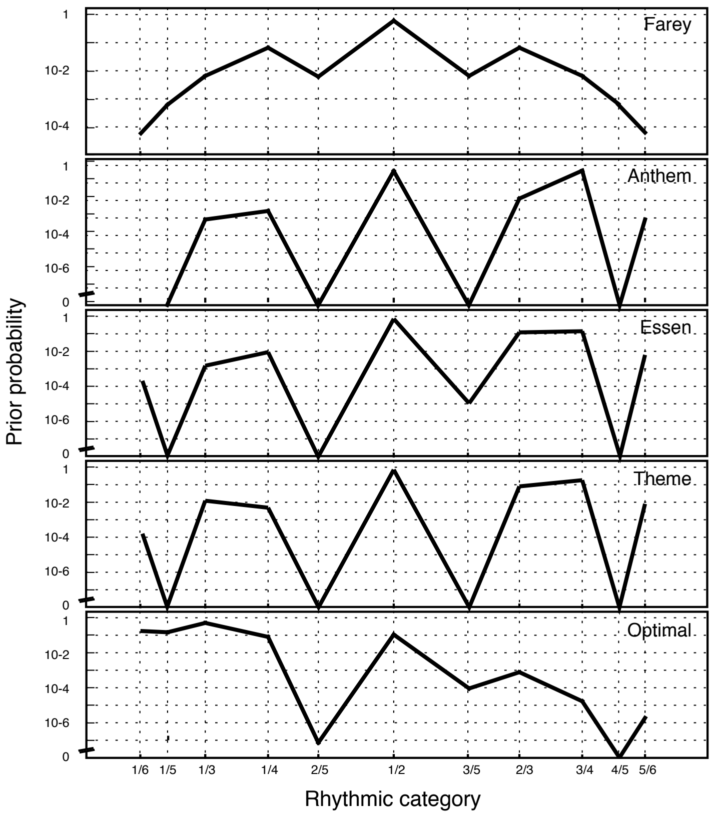


Figure 4.7 The optimal priors of the rhythmic categories obtained from Repp-Sadakata, the priors from candidate theories regarding the simplicity of ratios of temporal patterns represented on logarithmic scale and the priors as derived from a Farey tree on logarithmic scale. In the case of a zero prior, the corresponding value in the logarithmic graph was set to be the lowest rank.

Nevertheless, all priors appear to have in common a zigzag pattern. All kinds of priors agree in assigning a smaller value at the patterns having 5 as their denominators; at least they are smaller than the categories right next to them. The Farey tree is number-theoretic and not perceptually inspired, which is, for example, reflected in its 1/5 ratio—less complex than 1/6. The last ratio is usually considered to be perceptually simpler because it decomposes into a hierarchical duple plus triple subdivision. However, the score counts and the Farey tree, when used as priors, are found to give good predictions and share some characteristics with optimal priors. One observation of interest is well known from other empirical studies: the characteristics that exist in the priors derived from an empirical source (score count and optimal prior) and from a theoretical one (Farey tree). Quintuplets—rhythmic categories with ratios whose denominator is 5—are often mentioned as somewhat unstable patterns to produce or perceive. This instability has been associated with the nature of the mental coding of these temporal patterns (Povel, 1981; Povel & Essens, 1985).

The other artificial nature of the Farey tree is its perfect symmetry, but the assumption that rhythmic patterns retain their characteristics when reversed in time is not realistic. In contrast, Figure 4.7 shows that the structures of the optimal priors and s core priors are asymmetrical, as optimal priors tend to yield lower priors for long-short than for short-long patterns. The score counts also revealed asymmetrical characteristics but in the opposite direction, as long-short patterns tend to occur more often in musical scores than their short-long reflections. The asymmetry in processing of temporal patterns is also common in empirical studies. For production, Repp et al. (2002) found that 1/3 seems more difficult to perform than 2/3, as the number of trials that participants needed for a good performance according to their own standards is larger for 1/3 than for 2/3. Asymmetry is also commonly found in rhythm perception studies. For example, Desain and Honing (2003) showed asymmetry in the size of permutated rhythmic categories for three interval temporal patterns. Asymmetry in time suggests that auditory perception theories may need to be radically different from visual perception theories in which the symmetry in space plays a strong role. The different gradient between the optimal priors and the empirical score counts found here, however, is a puzzling phenomenon that we do not yet know how to interpret.

Discussion

By using Bayes' rule, we have found that rhythm perception data can be predicted accurately from rhythm production data and have concluded that the characteristics of rhythm perception and production processes can be successfully related[6].

The weakness of the many curve-fitting studies with free parameters—that they do not reveal anything about the flexibility of the theory or the likelihood of other outcomes—has been pointed out by Roberts and Pashler (2000). On this point, it is not surprising that optimized priors yield a good prediction because of the large number of parameters. Indeed, in general (discussed in Desain, Honing, van Thienen, & Windsor, 1998), even a theory that produces a perfect fit to the empirical data is not evidence in itself: there could be alternative explanations that are equally likely. Nevertheless, the optimized priors tell us what the best obtainable fit is with any corpus, giving a baseline and allowing a comparison with potential competing theories. The real test of the method uses the priors from several candidate theories with out any free parameters, such as Farey tree and score counts. It is encouraging for the validity of this approach that good results were obtained with these priors and that the optimal priors themselves are not very dissimilar to them. Furthermore, their characteristics are in agreement with that found in other empirical studies.

What does this method's success mean in terms of mental processing? Should Bayes' rule be considered just a methodological adjustment that makes possible the compensation for the effect of (nonuniform) competition in perception? By defining the strategy a perceiver can use when deciding which category a performance belongs to, we can formulate an answer. The optimal perceptual strategy, the one with the highest expected proportion of correct answers, maximizes the posterior likelihood (in the Bayesian sense) and chooses the rhythm with the highest probability, given the performance.

[6] It is more difficult to apply the procedure backwards. Given perception data and priors, production data are hard to constrain and predict, especially around the extremes.

Because human subjects behave close to this strategy, human rhythm perception is optimal, in the sense that it is adapted to, and optimized for, recognition in an environment in which rhythm production takes place—a result that may seem as trivial as it is deep. Progress in perception-action theories may eventually reveal the relevance of this optimality concept. Note that in our approach the production distribution must be fully known to the perceiver, a condition that may seem unrealistic. But advances in machine learning may guide us to a formal understanding of how this knowledge can be learned adaptively.

Implementation of Bayesian concepts for known results in perception and production of rhythm can shed new light and has consequences for theories about music cognition. In an example from the results of Sternberg et al. (1982), the conclusion could be drawn naively: even extraordinary, well-skilled musicians are not able reliably and accurately to produce and recognize time interval ratios in isolation, especially when they are more complex than 1/2. Furthermore, this work reported that rhythm perception and production are different for more complex ratios, and that performed ratios are far from their exact prototypes. The two processes were understood as not completely similar; they share only part of their mechanisms. But in the light of our Bayesian approach, one is drawn to the different conclusion that rhythm perception and production are closely associated; participants behave close to optimal in recognizing temporal patterns, even though the prototypes are far from exact. Furthermore, Bayes' rule allows us to attempt to understand the relation between the temporal processing of different patterns while using the concept of priors.

Further questions remain to be answered, such as what the optimal priors suggest and how prior knowledge is acquired. Training subjects on unfamiliar patterns or comparing rhythm perception, production, and score count data from different musical cultures may be the way to proceed toward a better understanding of the nature of these priors. Again in these cases we expect an optimal attunement of human perception to a world in which human production occurs, even with a changing world. Here individual differences may be modeled as a different set of priors. A more local adaptation may be required when rhythms are presented in the context of a meter or time signature. Because it is known that the "perceptive field," the area of sensitivity of a rhythmic category, is changed by metric priming (Desain & Honing, 2003), and certain rhythms are more probable in scores with a particular meter (Palmer & Krumhansl, 1990), it might be possible to predict contextual effects by changing the priors, as proposed by Friston (2002). This indicates that our method may eventually have consequences for the difficult cognitive modeling of psychological concepts including priming and attention.

Finally, one important issue not yet discussed is how this approach can be extended to handle more complex rhythms. Surely listeners do not memorize a huge number of distributions for different complex rhythms. Perception of complex rhythm must be based on simple rhythms in a principled way. Both Longuet-Higgins (1987) and Cemgil, Desain, and Kappen (2000) proposed this recursive metric subdivision; however, they assumed categories centered around mechanical timing. As we have shown here, even when one assumes mechanical performances, the perceptual categories may not align with them. A good model of how the perception of more complex rhythms can be derived from distributions of simple perceptual subdivisions is an open and difficult question.

Summary and Conclusion

In this study, we presented evidence that Bayes' rule can explain the relation between rhythm perception and production data by assuming that they are identical in a fundamental way. The validity of this approach was demonstrated, and consistent results were obtained under very different experimental conditions and computational setups. Using raw data sets, we simulated the way in which Bayes' rule relates the given probability distribution of the production of rhythmic patterns to the probabilities of the perception of rhythmic patterns.

Even with limited information of the data set, when only the means and standard deviations are known, it was possible to provide a relevant prediction.

Appendix 1 *Data description*

Desain & Honing (Perception)
In Desain & Honing (2003), a categorization experiment is described in which subjects respond with a rhythmic category that best reflects a three intervalperformance pattern, by using a computer interface for common music notation. More details, full results, and a model can be found in Sadakata & Desain (in preparation). In this experiment, 17 skilled musicians participated. Each stimulus pattern was made up of two time intervals on a time grid of 1/19 of 1000 ms, the minimum duration of an interval being three time grid units and the maximum 16 units. This yielded a set of 14 stimulus patterns. The pattern was repeated three times, embedded in a beat, as illustrated in Figure 4.3. The participants were asked to use notations commonly encountered in their practice. The set of possible response categories, using the computer notation interface, was still extremely large: thousands of ratios can be constructed using a range from whole-note to thirty-secondnote durations, using dotting, ties, triplets, etc. The actual responses used only 18 rhythms of the set.

Repp et al. (Production)
In Repp et al. (2002), the task was to perform simple monophonic melodies with the following rhythmic patterns: 1/2, 2/5, 1/3, 1/4, 3/5, 2/3, and 3/4. Twelve pianists participated and performed from a musical score. A maximum of three attempts was permitted without rehearsal, and the version that satisfied the performer was used. The rhythmic patterns were repeated over six bars and performed in four different tempi; a metronome was used before each performance. Averages over repetitions of inter-onset times were used. The responses are indicated as a probability density at the top of the left column in Figure 4.5.

Sadakata et al. (Production)
The task in Sadakata, Ohgushi, & Desain (2004) was to perform nine kinds of rhythmic patterns: 1/2, 1/3, 1/4, 1/5, 1/6, 2/3, 3/4, 4/5, and 5/6, in three tempo conditions and two playing m odes, mechanical and musical. The former was used in this study. Each pattern was performed from a score

and repeated 10 times (see Figure 4.3). Twelve percussionists participated in the experiments. Averages over repetitions of inter-onset times, excluding first and last bar, were used.

Sternberg et al. Experiment J2. (Perception)

In Sternberg et al. (1982), a number of perceptual experiments are presented. Three highly skilled musicians participated. In experiment J2, the subjects on each trial heard five beat clicks spaced 1000 ms apart, with marker clicks following the third and fourth, as shown in Figure 4.3. They were asked to judge the intervals from the beats to the markers. This beat-marker interval was varied over a range from a minimum of 43 ms to a maximum of 891 ms. They were presented as four different sets of 24 intervals whose spacing varied in the manner of a harmonic series. The participants selected a response from a set of eight categories, which are "less than 1/8 of a beat," "between 1/8 and 1/7," …, "between 1/3 and 1/2," and "greater than 1/2." The eight ordered categories define seven between-category boundaries on a hypothetical response continuum (see footnote 4). The estimated means of the psychometric function for each category and its variability were calculated. As the raw data of this study are no longer available, the means and variances were taken from Table 4.1 and Figure 4.5 of the original article.

Sternberg et al. Experiment P4. (Production)

On each trial in this experiment by Sternberg et al. (1982), 12 beat clicks were presented. The same musicians served as participants as in perceptual task (J2). Participants made 10 consecutive finger-tap productions to produce the ratio specified by instruction. The first response was produced after the third beat click, as illustrated in Figure 4.3. The ratio names used in the experiment were 1/8, 1/7, 1/6, 1/4, 1/2, 3/4, 5/6, and 7/8. Using the average value and standard deviations of these responses, the probability of occurrence of each category on a response continuum was calculated using a normal distribution. Averages were taken from Table 4.1, and variances were taken from Figure 4.5 of the original article.

Appendix 2 *Data description of a database of musical scores*

Barlow & Morgenstern, Dictionary of Musical Themes [Theme]

The *Dictionary of Musical Themes* (Barlow & Morgenstern, 1948, 1983) is a well-known theme index containing approximately 10,000 themes from the classical music repertoire. Both melody and rhythm are coded for each theme, as well as for its time signature. The collection consists of about 45% duple, 31% triple, and 24% compound meters.

Schaffrath, Essen Folksong Collection [Essen]

The *Essen Folksong Collection* (Schaffrath, 1993, 1995) contains a large sample of European folk songs, collected and encoded in the format of Essen Associative Code. Presently, 6,251 folk songs are available, although the total number of folk songs in the collection has reached 20,000. The metrical structure of the music (as signified by the time signature) is quite varied: 54% duple, 29% triple, and 17% compound meters. The database has been widely used to test a variety of music theories (e.g., Huron, 1999). Though mostly traditional German folk songs, they have simple rhythmic and metric structure and while regionally restricted, some songs are widely known in Europe (Huron, 2002). Thus it could be considered a reasonable sample of childhood exposure to music.

Shaw & Coleman, National Anthems Collection [Anthem]

The National Anthems Collection (based on Shaw & Coleman, 1960) is a corpus of the *National Anthems of the World* constructed for the evaluation of beat and meter induction models (Desain & Honing, 1999). The database contains only temporal information (rhythm and meter, no melodic or other information). The set (N = 105) consists of around 90% duple (70% is in 4/4 meter) and 10% triple meters.

Appendix 3 *Beta fit*

A beta distribution was chosen for the data approximation because of its flexibility, as it can fit skewed distributions. Furthermore, it has no tails extending to infinity. This is an advantage in the next step as minuscule differences in a set of small long tails (as, e.g., a normal distribution would exhibit) may result, after application of Bayes rule, in very large differences from the category center. This may generate noncontiguous categories. The beta distribution has two free parameters, α and β, that characterize the form of curve and two extra parameters (*w* and *m*) to rescale (squeeze) and shift the distribution to any mean and width using a linear transformation. In our study, the parameters α, β, *w*, and *m* were estimated from the production data set for every rhythmic pattern, using the maximum likelihood method. An example of the fitted beta distribution (and a normal fit to the same observations) appears in Figure 4.8. The vertical axis shows probability density, and the horizontal axis shows the duration of the first interval.

The production data for each ratio was thus characterized by the four parameters of the shifted beta distribution, yielding a family of curves illustrated in Figure 4.9, which also shows histograms of the raw data.

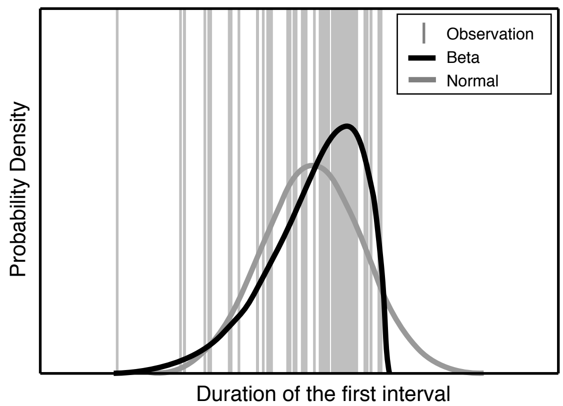


Figure 4.8 Example of the relation between observed data and two approximated continuous distributions. Gray vertical lines show the observations, the gray curve represents the data as approximated by the normal distribution and the black curve represents the data as approximated by a Beta distribution.

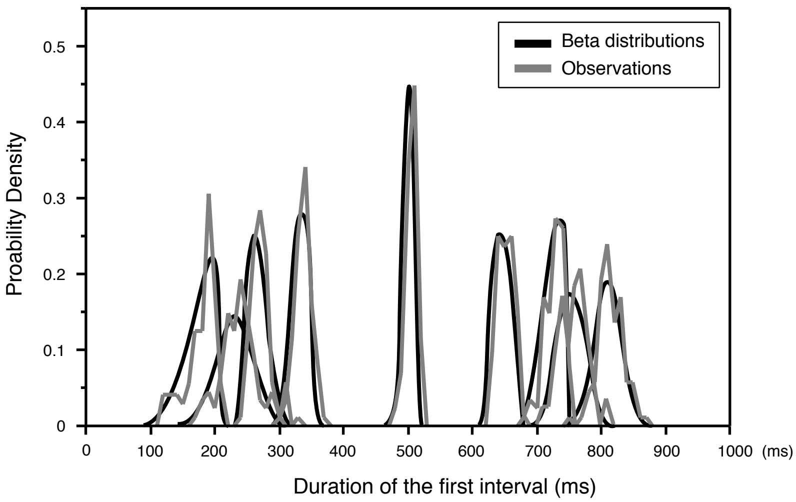


Figure 4.9 An illustration of the relation between the histogram of the observed production data (Sadakata et al., gray line) and the approximating Beta distributions (black line), for nine different rhythmic ratios.

Chapter 5

Diversity and commonality in rhythm perception and production: the influence of rhythmic complexity and culture

Abstract

In this study, two complexity measures of musical rhythm were examined in relation to cross-cultural (Dutch-Japanese) differences in rhythm processing by musicians. The measures were syncopation, which represents hierarchical complexity, and the durational variability of successive intervals, which represents serial complexity. Rhythm processing was assessed via perception, production and familiarity judgments, and these measures were shown to be related by Bayes rule, and explain the different findings in the different tasks. Cross-cultural differences emerged only for patterns that were complex in terms of their hierarchical structure: Dutch pianists produced shorter final intervals than Japanese pianists. This difference can be linked with similar findings for speech, and exposure to the rhythm of the mother tongue is a plausible explanation.

Possible musical universals have been proposed several times in the past, and indeed, in empirical studies some of measures of musical behavior have been shown to be independent of culture (Carterette & Kendall, 1999; Dowling & Harwood, 1986; Harwood, 1976; Sloboda, 1985). According to Drake & Bertrand (2001), one of the candidate universals for the processing of temporal patterns is a predisposition to reduce complex durational ratios simple ratios. As a consequence of this tendency, the stability or accuracy of rhythmic tasks decreases considerably when complex rhythm patterns are involved (Collier & Wright, 1995; Drake & Palmer, 1993; Peper, Beek, & van Wieringen, 1995; Povel, 1981). Fraisse (1956) even claimed that there are only two types of spontaneous inner representation of rhythm, namely short-long & long-short patterns having interval durations with a ratio of around 2:1 (see also Clarke, 1999). In theoretical accounts of this phenomenon, various authors take different approaches. Some are based on a hierarchical understanding of the temporal patterns and others apply Gestalt-like principles. However, all views address the heart of the issue that humans seem to be less able to handle inner representations of temporal patterns of increasing complexity. Contrary to the universality hypothesis, several recent studies have demonstrated non-universal behavior in rhythm perception and production in a real musical context (Drake & El Heni, 2003; Hannon & Trehub, 2005; Iversen, Patel, & Ohgushi, 2004; Ohgushi, 2002; Sadakata, Ohgushi, & Desain, 2004). The current study aims at investigating the commonality (musical universal) and diversity (non-universal) in relation to the complexity of the musical structure; We specifically studied the processing of temporal patterns comprehensively by looking at rhythm perception, rhythm production, familiarity judgemnets, and relating these results using the Bayes rule.

Cognitive theories of temporal processing are based on various features of rhythm, such as the accent structure of the patterns involved (Povel, 1984; Povel & Okkerman, 1981), number of the new subsequences encountered as the sequence develops (Lempel & Ziv, 1976), or one based on a purely mathematical notion of the complexity of rhythmic ratios, such as the Farey tree (Cvitanović, Shraiman, & Soderberg, 1985; Gonzalez & Piro, 1985). In particular, the processing of temporal patterns in musical rhythm has mostly been considered hierarchical (Martin, 1972; Povel, 1981). Here the use of the beat, which is subdivided or concatenated, is crucial to construct hierarchical levels. Thus, beat induction is one of the most fundamental issues for understanding processing of musical rhythm. It has been modeled in various ways, such as using a rule-based approach (Longuet-Higgins & Lee, 1982), optimizing the fit to an internal periodic clock (Essens & Povel, 1985), or postulating an internal oscillator coupled to the input rhythm (Large & Jones, 1999). In perceiving rhythm, once a beat is induced, a metrical structure can arise as the grouping of a sequence of a few beats. This metrical structure becomes the mental framework against which various aspects of the rhythm are evaluated, e.g., the strength of the beat according to its metrical position. Likewise, a syncopation, an unexpected missing note at an important metrical position (Lerdahl & Jackendoff, 1983), can only exist if a meter has been induced.

Several models have been able to provide a measure of the complexity of rhythm. For instance, Longuet-Higgins and Lee (1984) proposed a measure of syncopation strength that is based on the metrical strength of the missing note(s) calculated from the rhythm's hierarchical structure and note length. Another measure of sycopation, based on the efficiency with which a temporal patterns is proposed by Shmulevich and Povel (2000).

The fact that humans can also perceive and produce non-hierarchically structured temporal sequences, or at least less hierarchical structures, such as speech rhythm, raises the question of whether the process of rhythm perception and production is indeed essentially hierarchical. Demany and Semal (2002) showed that the ability of detecting regularity considerably decreases when the target temporal sequence contains isochronous intervals of which the duration was subdivided into two intervals following a random ratio. Thus, hierarchical cognitive processing of rhythm may be possible only when the predictable markers in the target sequence are salient enough, and we may need to look for surface level theories to describe regularity when that is not the case.

The normalized Pairwise Variability Index (nPVI) has devised to capture a serial feature of the temporal patterns, namely, the durational variability of successive speech units. It successfully accounted for the characteristics of speech rhythm of different languages (Grabe, Post, & Watson, 1999; Ling, Grabe, & Nolan, 2000). When a sequence of m intervals with durations d_1 d_2 $\ldots d_m$ is given, the nPVI is calculated as

$$nPVI(d_1 d_2 ... d_m) = \frac{100}{m-1} \times \sum_{k=1}^{m-1} \left| \frac{(d_k - d_{k+1})}{(\frac{d_k + d_{k+1}}{2})} \right| \qquad (1).$$

In this formula, the differences of all pairs subsequent durations, d_k and d_{k+1} are normalized by their means. The nPVI is lower when it is applied to a sequence that is close to isochronous, and is higher with a sequence of alternating long-short durations. Taking ds to represent intervals between vowel onsets in speech, the nPVI of so-called stress-timed language was found to be around 55-65, while that of so-called syllable-timed languages was around 30-45, and that of Japanese (a mora-timed language) was 41 (Grabe & Low, 2002).

Taking ds to stand for durations of notes in musical scores, Patel and Daniele (2003a) demonstrated that the successive durational variability of the ntoes is related to mother tongue of the composers: The musical nPVI (score nPVI) was higher for composers speaking stress-timed languages than for composers speaking syllable-timed languages. This finding was supported further with counts from various music databases (Huron & Ollen, 2003; Patel & Daniele, 2003b; Sadakata & Desain, submitted-a). These studies support the hypothesis that humans are sensitive to surface level, serial durational contrasts, during musical rhythm processing.

Another concern of the present study is the possible effect of cultural experience on rhythmic behaviors. When the complexity of the rhythm in an experimental task increases, a more non-universal musical behavior may be observed. For example, Drake & El Heni (2003) considered perceptual grouping in popular music. They studied how Tunisian and French musicians and non-musicians tap along with Tunisian and French popular music, and found that tapping rates of the participants synchronized at higher hierarchical levels with music from their own culture than with an unfamiliar type of music. This ability to synchronize at a higher hierarchical level may indicate a better understanding of the musical structure (Drake, Penel, & Bigand, 2000a, 2000b).

Hannon and Trehub (2005) demonstrated that North American adults performed more poorly on detecting alterations in an unfamiliar complex metre than in a familiar simple metre. On the other hand, Bulgarian and Macedonian adults, who are familiar with both simple and complex meters, and North

American infants, who have limited musical exposure, performed equally well in both conditions.

In yet another study, Iversen et al. (2004) showed a differential perceptual grouping by North American and Japanese adults of long-short cyclic temporal patterns. North Americans were inclined to group the temporal events into short-long patterns, while Japanese tended to group them into long-short patterns.

In rhythm production, the performance timing of specific rhythmic patterns in Mozart piano sonata was also found to be different between Japanese and Western professional pianists (Ohgushi, 2002). This finding was confirmed in a simpler musical context in which Dutch and Japanese percussionists were found to produce two-interval rhythms with different timing profiles (Sadakata et al., 2004). This effect only became significant for more contrasted durational ratios, such as 1:4 and 1:5.

These empirical studies may suggest that cultural differences in of the processing of temporal patterns occur as a result of exposure to a certain kind of music. However, this may not be a valid conclusion because, e.g., Japanese are exposed to western music almost as much as Westerners are. A more promissing hypothesis is that the results may reflect an effect of exposure to non-musical temporal patterns, such as speech. Therefore, it may be fruitful to look for similar differences in speech patterns to account for these culture-specific characteristics in processing musical rhythms. Indeed, some empirical studies have suggested that speech rhythm plays a role in the temporal processing of music (Huron & Ollen, 2003; Iversen et al., 2004; Patel & Daniele, 2003a; Sadakata & Desain, submitted-a).

In the present study, the syncopation rule as stated by Longuet-Higgins and Lee (1984) was applied as an indicator of hierarchical complexity, while the nPVI was applied as an index of serial complexity. We investigated the validity of these measures, as they are possibly probing the mental coding and representation by studying perception, production, and familiarity rating of the rhythms of which complexity was defined by the two measures. Furthermore, we examined whether the cultural background plays a role in determining the variations in observations.

Often rhythm perception and production data have been studied separately, and do not always agree (Sternberg, Knoll, & Zukofsky, 1982). Although many studies have related rhythm perception and production directly by comparing the means of the data (Drake, 1993b; Penel & Drake, 1998; Sternberg et al., 1982), this can be misleading (Desain & Sadakata, in preparation-b). In perception tasks, the stimulus set usually does not cover all the extreme areas of rhythm space, thus, responses in the categories at the borders of the space are truncated. Therefore, it is questionable whether means of perceptual categories are meaningful at all, especially those at the extremes. Besides, the nature of the tasks in the two types of experiment is very different. In the perception task, mental representations are in competition, because several categories can be associated with a target stimulus, whereas there is only one activated representation of the rhythm in the case of the production task. A Bayesian approach, which has been successfully applied to pitch perception model (Temperley, 2002) and rhythm transcription (Cemgil, 2000), offers an elegant model of (un-equal) competition among categories and makes it possible to reconcile relating data from these different tasks. Sadakata et al. (2006) have applied Bayes' formula in this way to show that the data distribution of perception results could be predicted well from the production distribution and the prior probabilities of musical scores.

Three experiments on processing three-interval rhythmic patterns were conducted. In Experiment 1, we compared two measures of rhythmic complexity with regard to how well they account for the perception of temporal patterns. The methodology was similar to that used by Desain and Honing (2003), in which participants notate the rhythmic category of presented sound stimuli. However, in the present study we presented a restricted set of common musical notations (scores) as response categories instead of allowing free transcription. In Experiment 2, we investigated the effect of rhythm complexity on the production of the same three-interval rhythmic patterns. In Experiment 3, we studied the participants' ratings of familiarity with the musical rhythms and compared them with the data obtained in Experiments 1 and 2. The results were also compared with a frequency count of the rhythms in a music corpus (Barlow & Morgenstern, 1948, 1983). Finally, the results from all three experiments were tied together using Bayes formula.

Figure 5.1 shows the relation between stimuli and responses. The term score refers to the rhythmic pattern in common musical notation, while a performance may contain deviations from the relative interval durations that

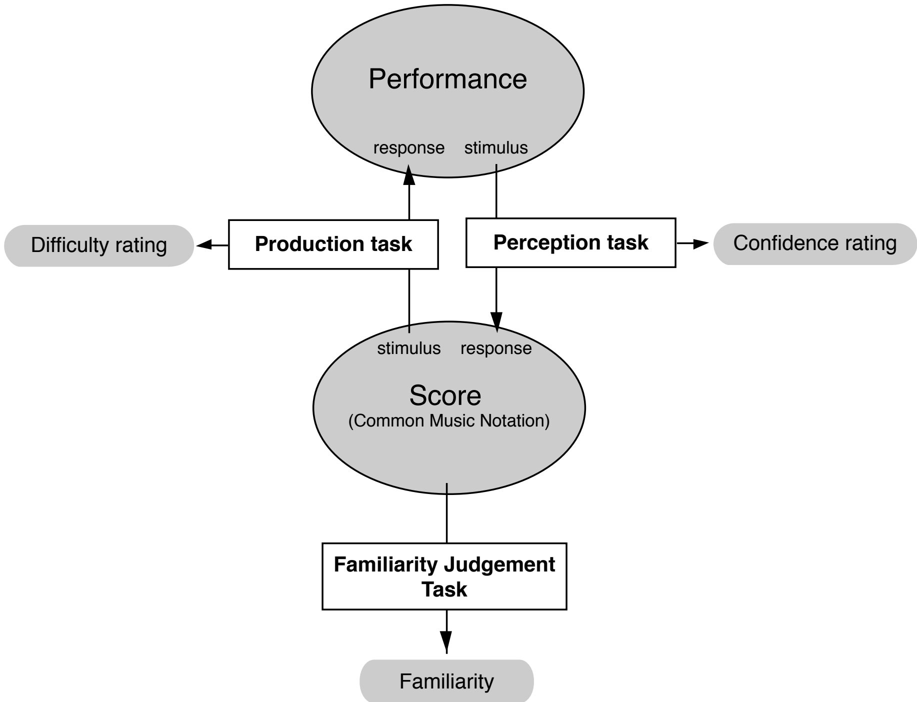


Figure 5.1 Relationships among the experimental tasks.

score specifies. In perception task, a performance served as stimulus and scores served as response choices; conversely, the score served as stimulus and the performance as response in the production task. Confidence and difficulty ratings of the tasks were collected from the perception and production tasks, respectively. Familiarity ratings were obtained with the score as stimulus.

The first goal of our study was to relate measures of cognitive complexity of the rhythms to the diversity of the results from rhythm perception, production, and familiarity rating experiments. The basic idea is that the simple patterns elicit consistent responses for every participant, whereas complex patterns induce more diverse responses. We want to find out which underlying mental representation (hierarchical or serial) is more salient by studying how well the two complexity measures account for the pattern of diversity in the experimental results.

Our second goal was to investigate a possible cultural impact on perception and production of complex temporal patterns. Although various cultural differences have been found in rhythm perception, there is still a lack of direct evidence demonstrating that listeners from different cultures systematically perceive the same rhythmic pattern in different ways. Experiment 1 examined this issue. In the production task of Experiment 2, we expected to find a similar cultural impact on three-intervals as on two-interval durational ratios (Sadakata et al., 2004).

Finally, our third goal was to determine the degree of consistency among the findings from our experiments. We expected similar results for perception and production experiments, because we designed the setup and stimuli to be equivalent for these experiments, and because the same participants took part in all the experiments. However, the similarity may only become evident when Bays rule is applied because as in perception, the familiarity or exposure (prior) becomes important while in production this may not necessarily be the case (see Sadakata et al. 2006).

Before explaining the experimental setups, the design of the rhythm set used in all experiments will be described. Three-interval patterns were sampled systematically from a performance space (Desain & Honing, 2003), comprising four sound onsets with a total interval duration of one second. The three axes in Figure 5.2a represent the three inter-onset intervals. Any rendition of a three-interval rhythm, for which the total duration is one second, is expressed as a point on the grey triangle surface, presented as a ternary plot in Figure 5.2b. The 66 rhythms used were sampled from this surface (dots on Figure 5.2b). To provide a resolution high enough to consistently probe the rhythm space, and to investigate the shape and position of possible rhythmic categories, the sampling was non-uniform[1]: it was more dense at the boundaries of the space where we know that rhythmic categories are packed to be packed more closely (Desain & Honing, 2003). The black dots in Figure 5.2b represent the 21 rhythms that were used in a separate perceptual consistency test.

The musical scores were required to cover a large area and scatter well over the performance space; however, at the same time they should not be too many. Furthermore, overly complex notations were to be avoided. We designed a

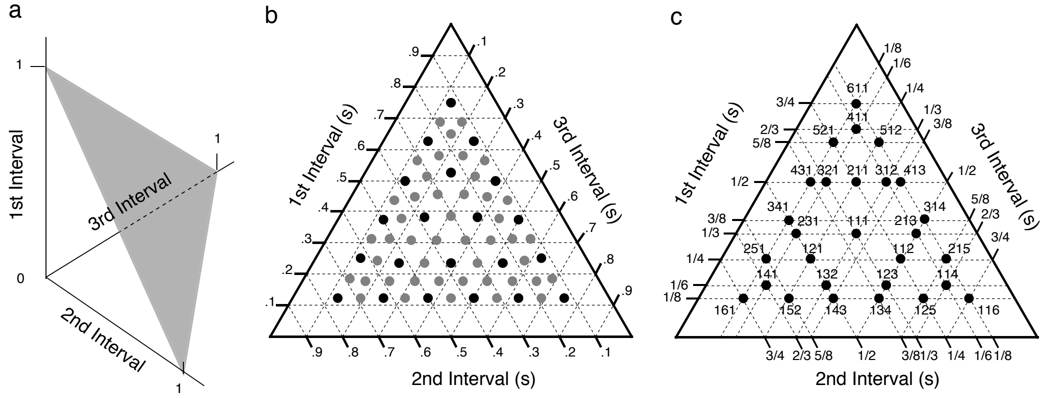


Figure 5.2 Rhythm space (a) and ternary plots (b, c) swhoing the rhythm stimuli (b) and scores (c).

systematic procedure to generate a reasonable set of score rhythms that meet these requirements. First, the three most common time signatures in Western tonal music were chosen (4/4, 3/4, 6/8). The eighth note was taken as the smallest interval in each meter condition, which is eight subdivisions of one musical bar in the case of 4/4 meter, and 6 subdivisions each for 3/4 and 6/8 meters. Next, the rhythms with no syncopation (syncopation is described in a later section) were listed in each time signature. This set consisted of 12 rhythms. Permutation of the durational ratios resulted in a set of 34 rhythms. In the case of a meter with a triple subdivision, the same durational patterns can be

[1] The performance space was sampled on a grid of 1/19th with a minimum time interval of 3/19th of a second, leading to triples of time intervals $[t_1, t_2, t_3]$ with $t_1 + t_2 + t_3 = 1$ and $t_i = n/19$ with $3 \leq n \leq 17$. To reduce the number of samples near the center of the space (around $t_1 = t_2 = t_3 = 1/3$) and allow a denser sampling close to patterns containing small time intervals, a correction *f* was applied, moving time intervals away from the center.

$$f(t_i) = t_i + c(t_i - 1/3)(g - min(t_1, t_2, t_3))/(g - 1/3)$$

With g is the grid size (1/19) c is the size of the correction (taken to be .5). This yielded a smallest time interval of .125 second.

represented in both 3/4 and 6/8 metric conditions. We used both time signatures in these cases. This brought the total number of possible three-interval rhythmic scores to 38 (18 scores for 4/4 meter and 10 each for 3/4 and 6/8 meters). This set covered the 21 most frequent categories (71% of all responses) in Desain and Honing (2003). The mapping of those scores into the performance space is shown in Figure 5.2c.

To classify the measure of this set, two complexity measures were applied. The hierarchical complexity measure as stated by Longuet-Higgins and Lee (1984), which defines the weight of notes according to their metrical position, is explained first. Pursuant to the rule, each metrical position was assigned place in a hierarchical tree structure, as shown in Figure 5.3. Metrical weight was defined according to the depth of the level in the metrical subdivision tree, counting down from 0. Syncopation occurs if a note onset following a metrically week note. In common music notation, this is represented as two tied notes with the second one in a stronger metrical position. Accordingly, syncopation was defined as follows. When N is the beginning note of the tie and R is a subsequent tied note, the strength of the syncopation is represented as the weight of R minus the weight of N (see Figure 5.3), where the 0 strength of the

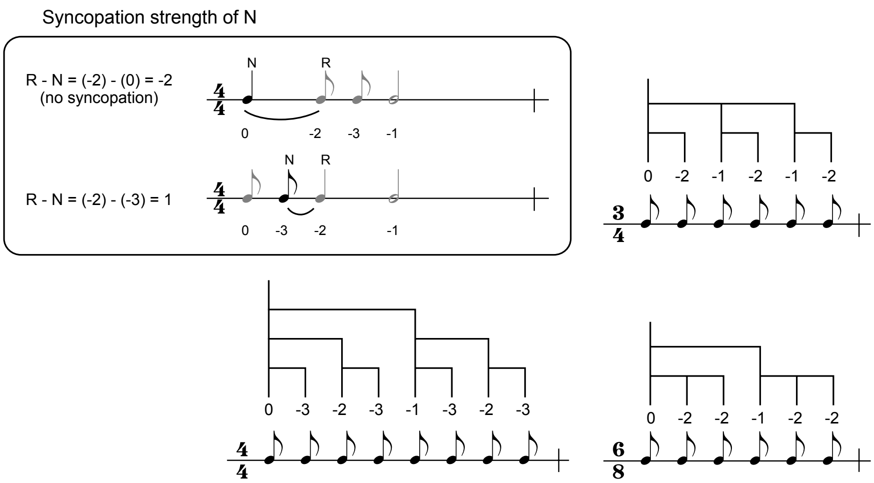


Figure 5.3 Weights of notes according to metrical position, and definition of syncopation strength as formulated by Longuet-Higgins & Lee (1984), applied to two example rhythms.

syncopation was defined as a marginal case. This syncopation rule reveals, for example, that the three-interval rhythm for which the durational ratio is 1-1-1 has a different syncopation strength in 3/4 meter and 6/8 meter. Syncopations in our three interval (four note) patterns can happen at two positions: in the second and the third note.

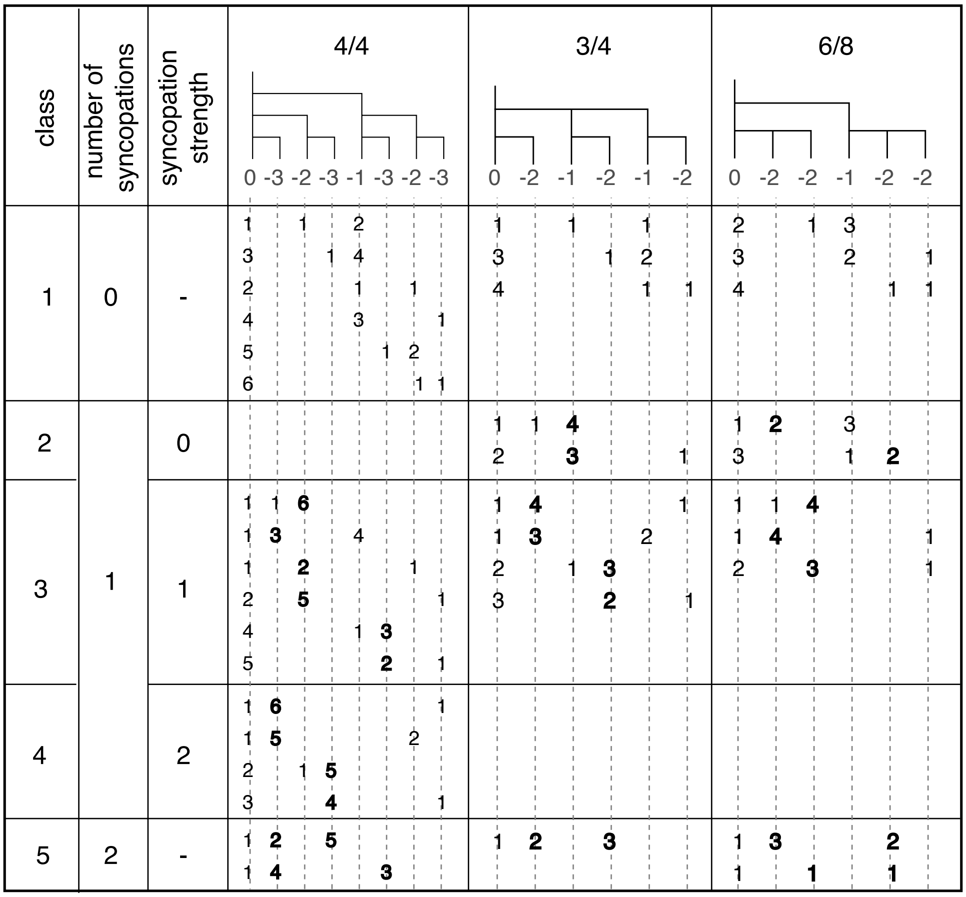


Figure 5.4 The rhythmic structure of the 38 scores. The numbers under the metrical tree indicate the weights of the notes according to Longuet-Higgins & Lee (1984). Each row represents a rhythmic pattern. The numbers indicates the relative intervale durations. Bold letters represent syncopated notes. The scores were classified into five levels of syncopation according to the syncopation rule.

Table 5.1 The syncopation level and the score nPVI for all the scores.

Time signature	Duration 1	Duration 2	Duration 3	Syncopation	nPVI
3/4	1	1	1	1	50
	1	1	4	2	76
	1	2	3	5	79
	1	3	2	3	96
	1	4	1	3	131
	2	1	3	3	83
	2	3	1	2	96
	3	1	2	1	83
	3	2	1	3	79
	4	1	1	1	76
4/4	1	1	2	1	63
	1	1	6	3	82
	1	2	1	3	93
	1	2	5	5	89
	1	3	4	3	88
	1	4	3	5	99
	1	5	2	4	124
	1	6	1	4	149
	2	1	1	1	63
	2	1	5	4	92
	2	5	1	3	124
	3	1	4	1	94
	3	4	1	4	99
	4	1	3	3	94
	4	3	1	1	88
	5	1	2	1	92
	5	2	1	3	89
	6	1	1	1	82
6/8	1	1	1	5	50
	1	1	4	3	76
	1	2	3	2	79
	1	3	2	5	96
	1	4	1	3	131
	2	1	3	1	83
	2	3	1	3	96
	3	1	2	2	83
	3	2	1	1	79
	4	1	1	1	76

For grouping the rhythmic scores, syncopation levels were defined as follows, 1: no syncopation, 2: strength 0 (marginal), 3: strength 1, 4: strength 2, and 5: 2 syncopations (see Figure 5.4). We assumed that the higher the syncopation level was, the more complex the rhythm would be.

The other measure, the score nPVI, was calculated for each score (see formula 1). To represent a serial nature of the rhythmic patterns, the nPVI was calculated in a cyclic manner taking the fourth intervals into account. For example, the nPVI for the sequence 1-1-1 was calculated as

$$nPVI = \frac{100}{4} \times \left[\left| \frac{1-1}{\frac{1+1}{2}} \right| + \left| \frac{1-1}{\frac{1+1}{2}} \right| + \left| \frac{1-3}{\frac{1+3}{2}} \right| + \left| \frac{3-1}{\frac{3+1}{2}} \right| \right] = 50 \qquad (2).$$

The complexity levels of the 38 rhythms according to these measures are shown in Table 5.1, revealing the contrast between the two measures. For example, because the nPVI captures only surface level fluctuation, it does not distinguish the same rhythm pattern presented in different time signatures, while the syncopation level does.

Experiment 1 Perception

Participants

Thirty-six paid volunteers took part in Experiment 1. Half of the participants were Dutch and the other half were Japanese. All were conservatory (classical) piano major students. The mean age of Dutch pianists was 22.8 years (ranging from 19 to 27) and that of Japanese pianists was 22.1 years (ranging from 20 to 27).

Materials

The stimuli were 66 performances of temporal patterns consisting of three intervals with a total duration of one second (see Figure 5.2b). The patterns were embedded in a metric context that was a sequence of one-second time intervals, which induced the bar-level of a meter, as illustrated in Table 5.5. We used the “high conga” percussion sound on a General MIDI synthesizer for presenting the stimulus patterns and the “low conga” percussion sound for presenting the bar markers. The responses choices were 38 scores. The perceptual experiment consisted of two parts: a perception test and a consistency test. For the consistency test, a subset of the stimuli from the perception test was presented again. The subset contained 21 rhythms that covered the whole performance space as shown in Figure 5.2b.

Procedure

The sound stimuli were presented through loudspeakers. The participants could adjust the loudness of the stimuli to a comfortable listening level. The sound presentation was driven by the POCO system (Desain & Honing, 1992; Honing, 1990), running on an Apple Macintosh Powerbook G4 and an iBook G3.

A specially designed computer interface (see Figure 5.6) was used. On the display, clicking the “play” button with the mouse allowed stimuli to be presented cyclically as long as desired. Rhythms were listed in common music notations separately for different metric conditions. First, one of the three different time signatures had to be chosen (3/4, 4/4, and 6/8). Several musical scores were presented with each meter option. After the selection of the score, the sound stimuli could be halted by the Stop-button. This choice of the meters and scores could be changed as many times as a participant wanted until the “Stop” button was pressed. Afterwards, participants were asked to indicate how confident they were about their response on a five-point scale, from “very confident” to “very unsure”. The “Ok” button indicated that the participants confirmed their response. In consistency test, the confidence rating was omitted.

The participants thus performed an identification task in which they had to associate the presented stimulus with a specific score. The stimuli were presented in random order. Participants were given the opportunity to practice the task (3 times) in order to get used to the procedure. They were instructed to think of the stimulus as if a percussionist played it. Then, they had to choose the score they thought was most likely used by the percussionist. The task was self-paced, and including breaks it took participants from about 45 to 90 minutes to complete the experiment.

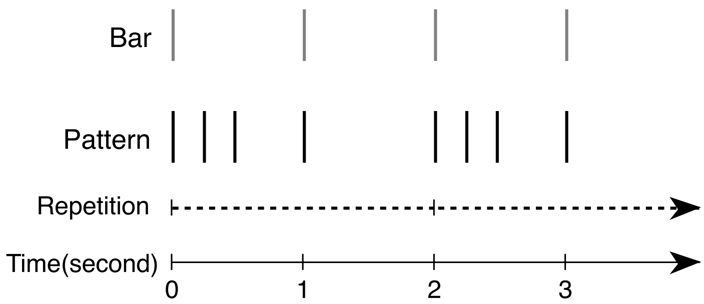


Figure 5.5 The stimulus pattern. Each line represents the onset of a percussive sound. The gray lines represent the metrical context against which the pattern that is to be identified (black lines) is repeated cyclically.

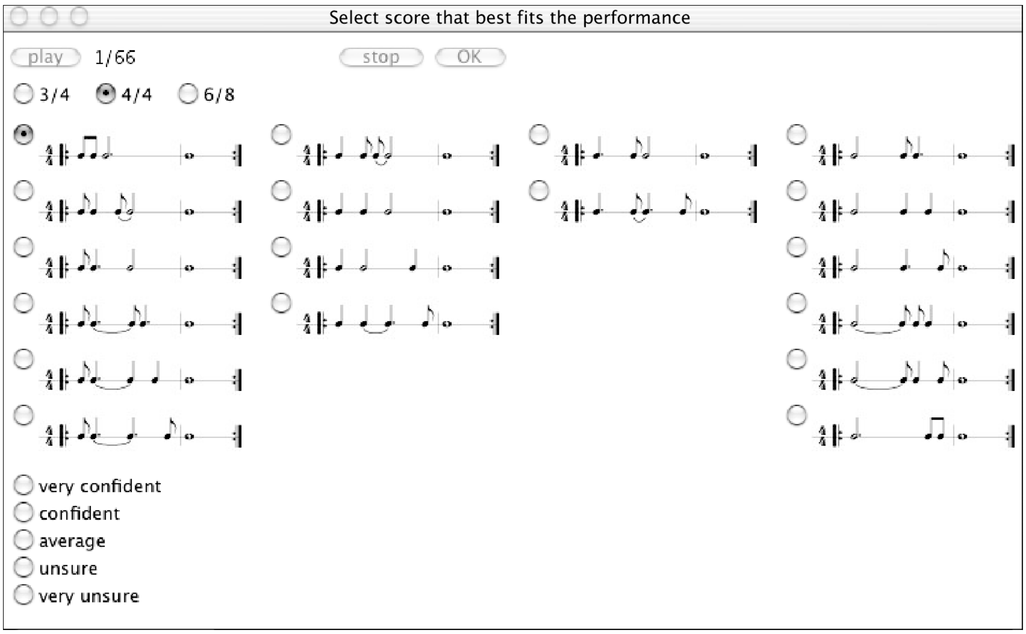


Figure 5.6 Music notation panel used as an interface for the identification task. The screen in which the subject has just selected a 4/4 time signature is shown.

Results

Perception test

Though the perceptual test yields score responses to each performance stimulus, this data is represented well by grouping and averaging the performance that gave rise to each response score. We refer to this as "perceived timing", this is the expected value of the performance timing given a specific score response. The perceived timings that had an extreme Maharanobis distance from the multivariate mean (centroid) of the IOIs were marked as outliers (4.5 %) and were removed from the data. The outlier analysis was conducted for each score condition by JMP (SAS, ver. 5.0.2). Next to the perceived timing of each score, standard deviation (SD) of the perceived timing between subjects, the number of the responses observed (response proportion), the number of the repetitions needed for the task, and the confidence rating were considered together with the syncopation level and the score nPVI. The confidence rating was normalized based on the observed response range of each participant. The performance nPVI and the tempo nPVI based on the perceived timing were also analyzed.

The correlation between data and complexity measures is shown in Table 5.2. The top tableau of Table 5.2 presents the correlations between the data of the experiments and the syncopation measures of the score (Spearman's Rho). The proportion of the responses correlated inversely with the syncopation level, revealing that the less syncopated scores tended to attract a larger number of responses. The syncopation level also correlated closely with the amount of repetition of the stimuli required for choosing the response, and did so inversely with the confidence rating. Thus, participants tended to require more time for choosing more syncopated scores and had a slight tendency to be more confident when they chose less syncopated patterns. Also, the SD of the second (SD2) and the third intervals (SD3) of the perceived timing correlated with the syncopation level. Thus, highly syncopated rhythms attracted more varied responses between participants.

In a more strict statistical examination of the relation between syncopation and the behavioral data, the between subject variability (SDs) of the perceived timing of the interval positions and the position of syncopated intervals (syncopation position) were compared. An ANOVA carried out on the SDs by interval position and syncopation position revealed a significant main effect of interval position ($F(2, 114) = 3.24$, $p < .05$), with SD2 slightly larger than the other two SDs. A main effect of syncopation position ($F(3, 114) = 7.89$ $p < .01$) showed that the SDs of the patterns having two syncopations in their second and third intervals or a syncopation in their third interval were larger than of the patterns having no syncopations. There was a significant interaction between these main factors ($F(6, 114) = 7.89$, $p < .01$). Further analysis (Tukey's HSD) revealed that the SD of the syncopated intervals was significantly larger than that of non-syncopated intervals. Thus, the perceived timing was more varied in the syncopated intervals than in the non-syncopated intervals. The syncopation level is a good overall predication of the perceptual results; it even predicts the position of more difficult (less acculturate) intervals within a pattern.

Among the given scores, 10 rhythmic patterns were represented in two ways, with the time signature 3/4 and 6/8. This provides us with an opportunity to systematically examine the role of time signatures. Because the performance stimuli did not contain any priming of the metrical context, the scores represented in a different metrical interpretation were only mentally imposed. Table 5.3 presents the mean perceived timing of three intervals in each time signature. It also shows the syncopation levels and the difference of the

Table 5.2 Correlations between perception data and complexity measures of the score (Spearman's Rho)

	Perception					
	Proportion	Confidence	Repetition	SD1	SD2	SD3
Syncopation level	-.42**	-.43**	.64**	-	.58**	.33*
nPVI	-	-	-	-	.37*	-
	Production					
	-	Difficulty	Repetition	SD1	SD2	SD3
Syncopation level	-	.74**	.67**	-	.68**	.43**
nPVI	-	.44**	-	.35*	.62**	-

Note. "Proportion" stands for proportions of responses attracted by the score (in Experiment 1). "Confidence" and "difficulty" refer to the ratings given by participants. "Repetition" means number of repetitions of the rhythmic pattern that participants needed for performing the task. The standard deviations of each interval are shown as SD1, SD2 and SD3 where each number corresponds to the interval location in the sequence. An asterisk (*) indicates the significance level of correlations (**= $p < .01$, *= $p < .05$). Non significant correlations are omitted.

Table 5.3 Mean durations of the three intervals for 10 rhythmic patterns given in 2 time signatures.

Rhythm	p	Syncopation level			3/4			6/8		
		3/4	6/8	distance	T1	T2	T3	T1	T2	T3
1-4-1	0.27	2	2	0	181.4	652.0	166.1	170.4	659.3	170.0
4-1-1	0.21	0	0	0	645.1	166.1	188.4	649.1	169.3	181.4
1-1-4	0.31	1	2	1	164.6	174.3	660.7	168.6	171.4	659.7
2-3-1	0.98	1	2	1	353.9	475.0	171.0	355.8	479.0	165.1
3-1-2	0.21	0	1	1	502.3	154.0	343.4	514.6	161.0	324.3
1-3-2	0.24	2	4	2	157.9	503.3	338.7	179.4	475.2	345.4
2-1-3	0.73	2	0	2	349.6	163.9	486.4	332.2	179.9	487.8
3-2-1	0.87	2	0	2	514.8	316.1	168.9	492.6	339.3	168.0
1-2-3	0.04*	4	1	3	169.4	330.8	499.7	150.2	370.6	478.9
1-1-1	<.001**	0	4	4	335.4	320.8	344.8	334.3	338.9	327.6

Note. The syncopation level and the difference of the syncopation level between patterns in different time signatures are also shown. An asterisk (*) indicates a significant difference of the durations between different time signatures ($p < .05$).

syncopation level between patterns when notated in different time signatures. A MANOVA carried out on the local tempo, the deviation of perceived timing from the score timing, by time signature for each rhythm pattern revealed a significant effect of the time signature on 1:1:1 ($F(1, 115) = 25.45$, $p < .01$) and 1:2:3 ($F(1, 133) = 4.26$, $p < .05$) patterns. Note that the 1:1:1 and 1:2:3 patterns exhibit the largest difference between the syncopation levels when given in the different time signatures.

Now we turn to a non-hierarchical measure. The score nPVI calculated from the score timing also served as a complexity index of the scores. No significant relation was found between the score nPVI and proportion of responses (see Table 5.2). Thus, predicted complex rhythms, with higher nPVI, were chosen as often as simple rhythms, with lower nPVI, as a response. As shown in the top tableau, second row in Table 5.1, there was a small but significant correlation between score nPVI and the SD of the second interval (SD2). Because the significant correlation between nPVI and SD2 might be due to the correlation (.45) of score nPVI and the syncopation level, we calculated the correlation of the score nPVI and SD2 partialing out the effect of syncopation: It dropped to .17. In contrast, the partial correlation of the syncopation level and SD2 stayed high at .56. Accordingly, this significant correlation between the score nPVI and SD2 is caused by the fact that they both correlate with the syncopation level.

The eight triangular plots shown in Figure 5.7 present the perception results for the two cultural groups. The gray surfaces correspond to the areas where a significant majority response largest proportion of response is significantly larger than the second largest response ($p < .01$) was observed and the white surfaces to the areas where there was no clear majority. The two triangles on top show the pooled result of all scores. Patterns similar to the time clumping map described in Desain and Honing (2003) appeared in the plots. The scores made up of lower integer ratios constituted a significant fraction of the total probability mass of the responses; the patterns that were based on the subdivision-unit of 3, 4 and 6 covered 67% of the total responses. In fact, the 10 patterns from the Dutch group and 9 patterns from the Japanese group concur with the 12 most frequent response scores in Desain and Honing's study. Although some patterns based on eight subdivision-units (3-1-4, 4-3-1) attracted a relatively large amount of the responses (3.8 %, 4.1 %, respectively), some of these patterns (1-4-3, 3-4-1) did not appear as response at all. Note that these patterns are all permutations of each other: consist of the same durational ratios, but are organized in a different order. Thus, an asymmetry on rhythmic categorization was observed, in agreement with Desain and Honing's findings.

The bottom six triangles in Figure 5.7 show the results separated for each time signature. Although category boundaries cannot be inferred from these plots, they allow us to make an easy comparison between the cultural groups and to easily identify the preferred time signature. A multivariate analysis of variance (MANOVA) carried out for the local tempo (deviation from the score duration) of perceived timing by the cultural groups for each score showed a cultural difference only on two patterns, 3/4-3-2-1 ($F(1, 15) = 8.26$, $p < .05$); 6/8-1-4-1 ($F(1, 75) = 6.26$, $p < .05$). Despite the cultural difference, there is no consistent general direction of the effect.

Consistency test

The results consist of the within-subject agreements calculated from the responses for each stimulus pattern presented four times (three times in the

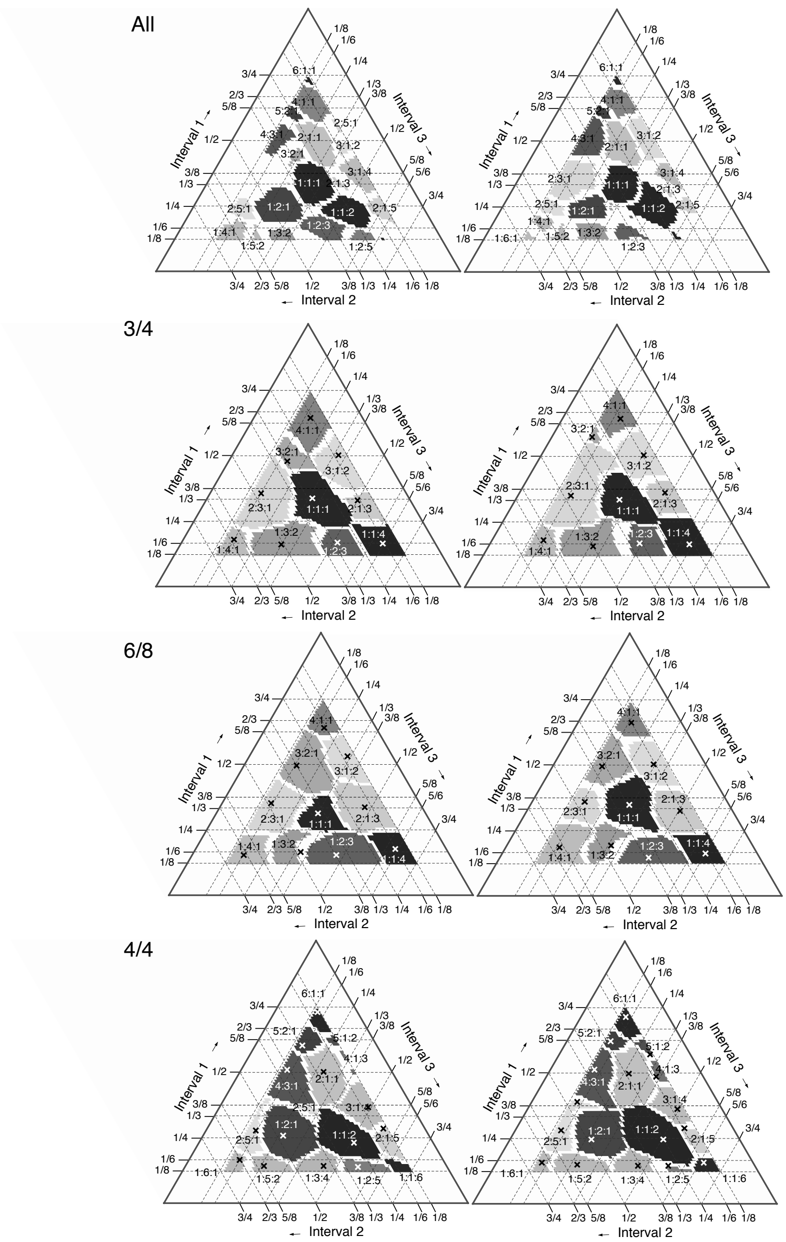


Figure 5.7 Perception results for the two cultural groups (Japanese on the left, Dutch on the right).

consistency experiment and one time in the perception experiment). The same rhythm patterns with different time signatures were counted as separate responses. The average agreement was 69 % for Dutch and 71 % for Japanese group. An ANOVA carried out on within-subject agreement by cultural group and the kind of the performance stimuli revealed a significant main effect of performance stimuli ($F (20, 756) = 20.26$, $p < .01$). There was no significant cultural difference ($F (1, 756) = 2.39$, $p = .12$) and no interaction between stimuli and cultural groups ($F (20, 756) = 0.98$, $p = .48$). It showed that the within-subject consistency was equivalent between cultural groups.

Discussion

A meaningful association between the amount of syncopation and the data was observed; for instance, perceptual timing was more varied in syncopated intervals. The syncopation level also accounted for the other data reasonably well, while the score nPVI hardly did. Thus, the hierarchical structure of the rhythm seems to play a part in the perceptual process much more than serial surface order. The finding that the perceived timing is significantly different when the same pattern was perceived in different hierarchical structures strongly supports this hypothesis.

The results were well in accordance with the perceptual study carried out by Desain and Honing (2003). As in that study, the scores made up of lower integer ratios constituted a significant fraction of total responses, and the asymmetry in the position and shape of rhythmic categories was found as well. These agreements suggest that providing limited scores for a response instead of allowing free notation did not influence the results of the categorization task.

Although the perceived timing of the two scores was significantly different between the cultural groups, examination of the perception results suggested no systematic cultural effect on the data. As such, the data failed to support our hypothesis that listeners from different cultures perceive the same rhythm pattern in different ratios. This is rather surprising, because many authors have repeatedly found a cultural difference in music perception, as described in the introduction. A possible explanation of our results relates to the participation of trained musicians. Most of the cross-cultural differences have been demonstrated with non-musicians, while our experiment involved only highly trained musicians. Thus, if there is any, musical training might cancel out culturally specific perception behavior. If so, in other words, the cultural characteristics could be unlearned after a certain amount of the training. Further research is therefore required for characterizing this process.

The results of the consistency test clarified two important issues. First, a relatively high consistency within individuals on the task (around 70 % on average) indicates that the participants were able to carry out these tasks reliably. Furthermore, as supported by the absence of significant cultural differences between within-subject consistencies, a high similarity between groups indicates that the participant groups were properly matched in terms of skill level.

Experiment 2 Production

Participants

The same 36 pianists from Experiment 1 served as participants in this experiment.

Apparatus

The scores were presented on the screen of the computer (Apple Macintosh Powerbook G4 or iBook G3). A MIDI keyboard (YAMAHA DX-7) connected to the computer was provided for the performance using General MIDI piano sounds. The participants could adjust the loudness of the sound to a comfortable listening level. Responses were recorded and metronome sounds were driven by the POCO system (Desain & Honing, 1992; Honing, 1990).

Stimulus materials

The same scores (38 rhythmic patterns) as in Experiment 1 were used as stimuli, but now to be performed (see Figure 5.4). Accordingly, the responses were the performance of these scores.

Procedure

A specially designed computer interface (see Figure 5.8) was used to present scores to the participants. Clicking the "Next" button with the mouse allowed them to start a cyclic metronome sound. After the performance, the metronome sound could be halted by the "Stop" button. Next, participants were asked to rate the difficulty of the task on a five-point scale, from “very difficult” to “very easy”. Response was confirmed by the “Ok” button.

The participants completed a production task in which they had to perform the presented scores on a keyboard along with the metronome sound six times. They were allowed to redo the task until they were satisfied. The presentation of the score and the responses were self-paced. The stimuli were presented in random order. Participants were given the opportunity to practice the task (3 times) before the experiment in order to get used to the procedure. Participants were allowed to redo each rhythm until they were satisfied with their performance. Including short breaks, the task took 15 to 30 minutes to complete.

Results

Four repetitions of three IOIs of the performances obtained under the given score stimuli were analyzed. The MIDI files were converted to tables of IOIs of the performance using POCO. The total durations of the three IOIs were

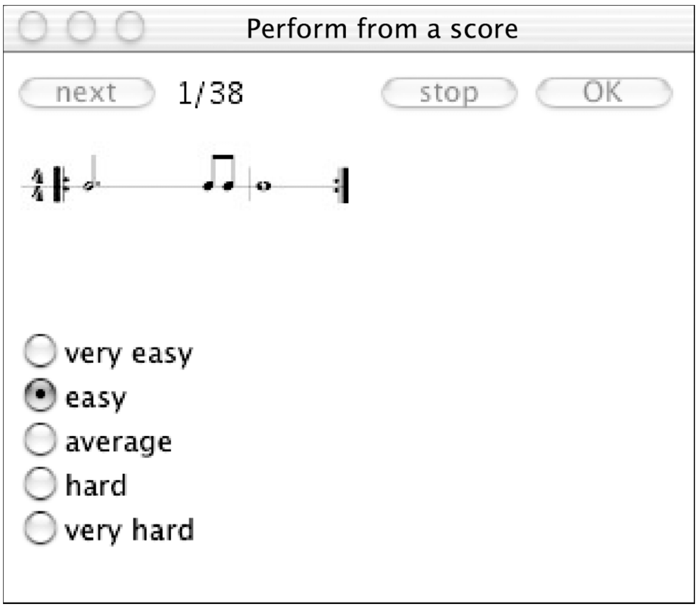


Figure 5.8 The display of a music score on the interface in the production task.

normalized to one second. Each performance of the production trials includes six repetitions of the patterns. We used an average of the four repetitions (excluding the first and the last repetition) for the analysis.[2] The averaged data having an extreme Maharanobis distance from the multivariate mean (centroid) of the IOIs were marked as outliers (1.6 %). The outlier analysis was conducted for each score condition by JMP (SAS, ver. 5.0.2). The consistency of the production (within-subject SD and between-subject SD), the repetition needed for achieving the performance with which they were satisfied, and the difficulty rating were considered in relation to the syncopation level and the score nPVI. The difficulty rating was normalized based on the observed response range for each participant. As for the perception results, the performance nPVIs and the tempo nPVIs calculated from the timing of the production were considered.

First, the contribution of the syncopation level is considered. Correlations between the production data and the syncopation level are shown in the first row of the second tableau in Table 5.2. It shows that the variability between participants of the performance of the first and second intervals (SD2 and the SD3) was closely correlated with syncopation level. Also, the difficulty rating was highly correlated with the syncopation level, as well as the number of repetitions needed to produce the ideal timing of the performance. A consistency measure of the production, as within-subject SD in timing of the production, was calculated from the four repetitions of the production. An ANOVA performed on the within-subject variability by syncopation level showed a significant main effect of syncopation level (F (4, 260), 5.61, $p<.01$). Further analysis (Tukey's HSD) revealed that the within-subject variability of the scores classified in syncopation level 5 was significantly higher than the other scores. Because the more syncopated patterns were more difficult to perform, the tasks involving these patterns required more repetition and yield more variety in the produced timing.

[2] Some responses included fewer than six usable repetitions. We distinguished repetitions that did not include three intervals due to, e.g., a miss hit on the keyboard, and of which the total duration lay outside the range of 500 and 1500 ms. Accordingly, we excluded performances that contained fewer than three repetitions from further analysis. When selected repetitions consisted of just four repetitions, we calculated the average using all repetitions. For performances including five repetitions the average timing was calculated from the second to fifth repetitions.

As in perception, an effect of the interval position and the position of syncopation intervals occurred (syncopation position) on between subject variability (SDs) was tested. An ANOVA on the SD of the timing by interval position (three positions) and by the syncopation position was conducted, showing a significant main effect of the syncopation position (Dutch; $F(3, 114) = 9.11$ $p < .01$, Japanese; $F(3, 114) = 18.34$ $p < .01$), as well as of interval position (Dutch, $F(2, 114) = 2.65$ $p < .01$; Japanese, $F(2, 114) = 3.6$, $p < .05$) and a interaction between these two factors (Dutch, $F(3, 114) = 5.87$ $p < .01$; Japanese, $F(6, 114) = 7.50$, $p < .01$). Further analysis (Tukey's HSD) showed that the SD of the syncopated intervals was significantly larger than that of non-syncopated intervals in both cultural groups, consistent with the perceptual result.

Significant correlations between the score nPVI and the data were observed as well (see the last row of the second tableau in Table 5.2), namely, with SD1, SD2, and with difficulty rating. However, because syncopation level and score nPVI are correlated, these correlations need not point to a separate explanatory power of score nPVI. Indeed, the partial correlations between score nPVI and

Table 5.4 Rhythms of which performed timing was significantly different between the cultural groups. The syncopation level, the score nPVI, and the results of the statistical analyses are shown.

Rhythm	Syncopation level	Score nPVI	Statistics
4/4-2-1-1	1	63	$F(1, 34) = 5.79$, $p < .05$
6/8-2-1-3	1	83	$F(1, 33) = 4.98$, $p < .05$
6/8-1-2-3	2	79	$F(1, 30) = 8.11$, $p < .01$
3/4-1-1-4	2	76	$F(1, 34) = 9.02$, $p < .01$
3/4-2-3-1	2	96	$F(1, 33) = 3.56$, $p = .07$
4/4-1-3-4	3	88	$F(1, 32) = 5.23$, $p < .05$
4/4-5-2-1	3	89	$F(1, 31) = 5.03$, $p < .05$
6/8-1-4-1	3	131	$F(1, 32) = 7.12$, $p < .01$
3/4-1-2-3	5	79	$F(1, 31) = 22.15$, $p < .01$
6/8-1-3-2	5	96	$F(1, 32) = 3.79$, $p = .06$
4/4-1-4-3	5	99	$F(1, 29) = 4.54$, $p < .05$
4/4-1-2-5	5	89	$F(1, 30) = 3.94$, $p = .06$

the data (removing the effect of syncopation level) dropped considerably (SD2 r=.40, difficulty rating r=.24), while that between syncopation level and data (excluding an effect of nPVI) stayed high (SD2 r = .60, difficulty rating r = .72). Thus, again, the nPVIs did not account for the production data as much as the syncopation level did.

The examination of the production results revealed a cultural difference on the production timing for the 12 scores, including a few scores that the difference in production timing almost reached significance. These rhythms are provided in Table 5.4 with the results of the statistical analyses, the syncopation level and the score nPVI. The number of the rhythms that were significantly different between cultures increased with the syncopation level, while there was no regularity observed with the score nPVI.

Figure 5.9 presents the local tempo change (deviation from the mechanical duration) of the second interval on the vertical axis, and that of the third interval on the horizontal axis. Each data point represents a performance of one score by a single participant. The cultural bias in local tempo change can be clearly seen in these scores that include two syncopations (syncopation level 5).

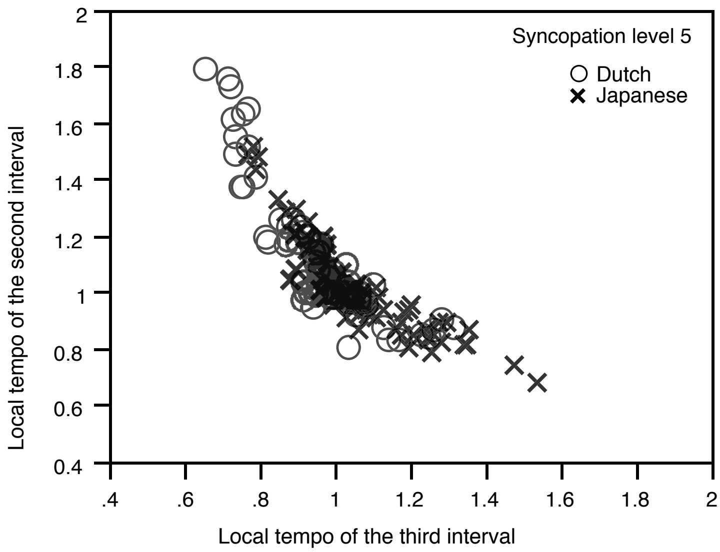


Figure 5.9 Deviation from exact score timing of the second and third intervals of the produced rhythms.

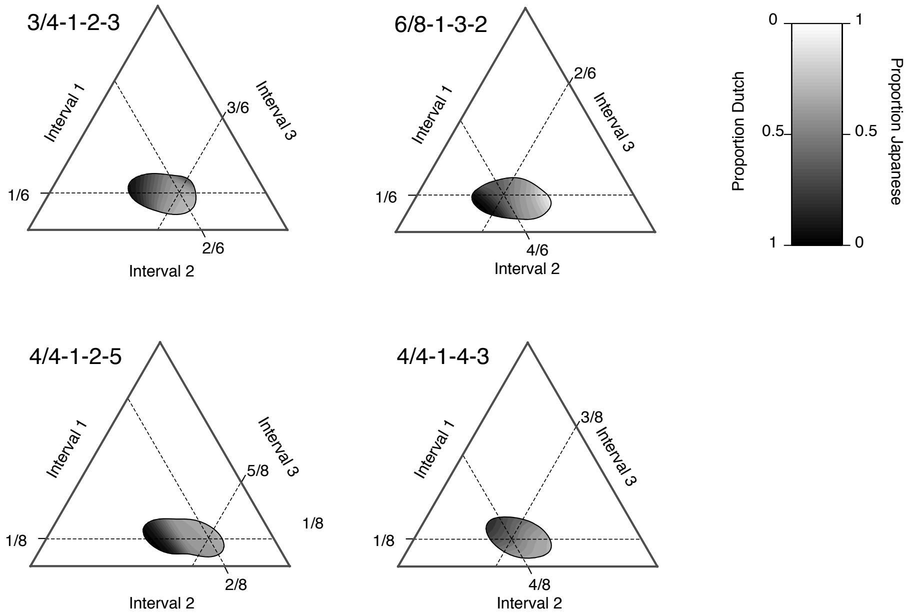


Figure 5.10 The data distribution of the four scores of the highest syncopation level plotted in a triangular plots. The cultural groups timed differently in production. The shapes in each triangle represent the area within that 95 % of the observations occurred. The shade corresponds to each culture: a white area indicates more observations for the Japanese group than for the Dutch group, and a black area indicates the opposite.

To check if all four scores showed the same tendencies, we plotted distributions of responses separately for each score using Parzen method (Parzen, 1962) in a performance space. The four triangle plots in Figure 5.10 show the scores for which the cultural groups produced significantly different timings. The boundary of the irregular shapes in each triangle encloses the major portion of performance responses (more than 90 %) for that score. The shade corresponds to the proportion of each culture. White indicates that more observations occurred for the Japanese group than for the Dutch group, black indicates more Dutch observations than Japanese ones, and gray color indicates an equal distribution of responses between cultures. The plots reveal an overall shift of Dutch responses towards the left corner of the triangular plot compared to the Japanese production. This shift indicates the prolongation of the second interval and a reduction of the third interval for the Dutch group relative to the Japanese group.

Discussion

Unlike the perceptual experiment, the score nPVI in this experiment showed significant correlations with some of the data, (i.e., SD1, SD2, and difficulty ratings). Although the partial correlations revealed that these correlations largely depend on the correlation of the nPVI with the syncopation level, it is reasonable that the more different successive durations are, the more difficult they are to perform.

There was a significant influence of syncopation on production. The results show parallels with the two-interval production study of Sadakata et al. (2004), in terms of the relation between variability in the timing of production and the degree of cognitive complexity of the rhythm, in a hierarchical sense. Both studies found more variability in produced timing for more complex rhythms, and that culture-specific tendencies in production timing only emerge in such complex rhythms.

Compared to the cultural difference found in the two-interval production study, the results with three-interval production were less straightforward. The expected culture-specific tendency, that is, for Japanese pianists to perform durational ratios in a less contrasted manner than Dutch pianists, was not observed. Instead, we found a systematic cultural distinction in a different form: The Dutch pianists were inclined to prolong the second interval compared to the Japanese pianists, and the Japanese tended to prolong the last interval compared to the Dutch pianists. These tendencies were complementary, because the total duration of the three intervals had to be constant. The tendency became apparent at the intervals where more variety in the performance timing was observed. Therefore, it became prominent when the scores included two syncopated intervals, as the SD of performance timing was significantly larger for syncopated intervals. Note that among the scores with the highest syncopation level, the only score that was equivalent between the cultural groups was 6/8-1-1-1. It is well known that the isochronous pattern is exceptional as compared to the other patterns. It has been shown to be a robust perceptual category independent from the cultural background and musical training (ten Hoopen, Sasaki, Nakajima, Remijn, Massier, Rhebergen, &

Holleman, in press). Participants may have locally switched meter to the triple subdivision (3/4-1-1-1), because of the strong evidence for that meter in the pattern.

In the study by Sadakata et al. (2004), the cultural difference found in two-interval patterns (Sadakata et al., 2004) depended on the durational ratio regardless of interval order. This is in contrast with the difference found in this study. Examples of patterns with two syncopations are 3/4-1-2-3 and 6/8-1-3-2, in which the second interval was prolonged consistently by Dutch participants and the third interval by Japanese participants. Considering the duration of the last two intervals, the two-interval study would have predicted them to be contrasted by the Dutch (the second interval longer in 1-3-2 and the third interval longer in 1-2-3) and assimilated by the Japanese (the second interval shorter in 1-3-2 and the third interval longer in 1-2-3). The incompatibility between the result for two-interval and three-interval patterns is in accordance with the study by Repp, Windsor, and Desain (2002), who showed a different assimilation tendency of durational ratio in the production of three-interval as compared with two-intervals. It suggests that the production timing of three intervals is not serially composable, that is to say, for example, the performance timing of 1-3-2 cannot be predicted from the combined timing of the 1-3 and 3-2 patterns. This is in accordance with the likely cognitive strategy that musical rhythms are processed in a hierarchical manner. Indeed, the large effect of syncopation level strongly supports this hypothesis.

Experiment 3 Familiarity and exposure

Participants

The same 36 pianists from Experiment 1 served as participants.

Materials

The same scores (38 rhythmic patterns) as in the other experiments were used as stimuli to be rated (see Figure 5.4). The frequency of occurrence of the same 38 scores was counted in The Dictionary of Musical Themes (Barlow & Morgenstern, 1948; 1983), a well-known theme index containing approximately 10,000 themes from the classical music repertoire.

Procedure

The scores were presented on the screen of the computer (Apple Macintosh Powerbook G4 or iBook G3). A specially designed computer interface (see Figure 5.11) was used in which the rhythms were shown in common music notation. Clicking the 'next' button with the mouse allowed a score stimulus to be presented. Participants were asked how common these scores were in their daily music practice using a five-point scale from "very common" to "very rare". The "Ok" button signaled that the participants had confirmed their response. The task was self-paced and took from 10 to 25 minutes to complete.

The counting of frequency of occurrence in the database was conducted in the same manner as in Sadakata, Desain, and Honing (2006) by using the POCO system. Only rhythms consisting of a bar with three notes were considered. The rhythms that exactly matched the ones used in the experiments (38 scores) accounted for 54.3 % of all three-interval rhythms in the database for which the total duration of intervals constitute a bar (2,650 rhythms). This rises to 84.0 % if only rhythms in 3/4, 6/8 and 4/4 meters were considered. Seven of the scores did not appear in the frequency counts (3/4-1-2-3, 3/4-3-2-1, 6/8-1-3-2, 4/4-1-4-3, 4/4-1-6-1, 4/4-2-1-5, 4/4-3-4-1).

Results

The familiarity rating was normalized on the basis of the observed response range of each participant. The mean ratings showed no cultural difference, ($t(76)=-1.16$, $p=0.25$), suggesting that overall familiarity with scores was comparable for the two groups.

A high correlation was obtained between familiarity rating and score count ($r=.82$, $p<.01$), which suggests that the familiarity ratings of the participants reflect exposure to Western tonal music. Alternatively, however, the correlation may reflect the fact that less complex rhythms are more frequent and seem more familiar.

The familiarity rating correlated with the proportion of responses in the perception experiment ($r=.57$, $p<.01$), indicating that the scores rated as more familiar tended to be selected more often. Furthermore, it was highly inversely correlated with the difficulty rating from the production experiment ($r=-.89$, $p<.01$). Thus, the scores rated as more familiar tended to be easier to produce.

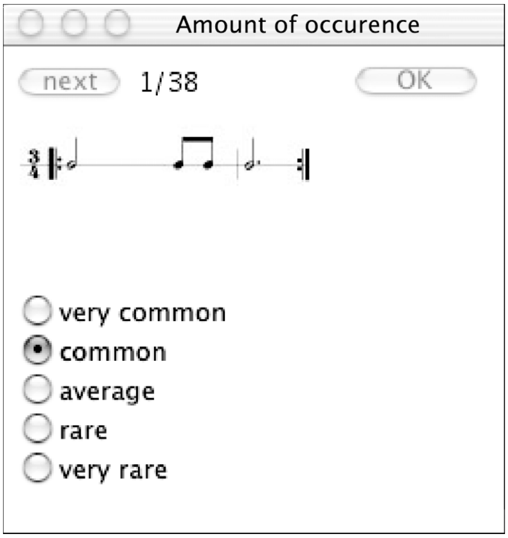


Figure 5.11 The display of a music score used as interface in the judgment task.

The two complexity measures (syncopation level and score nPVI) were inversely correlated with familiarity rating (syncopation level, $r=-.85$, $p<.01$; score nPVI, $r=-.57$, $p<.01$). Neither the partial correlation between the syncopation level and familiarity rating excluding the effect of the score nPVI, nor the one between the score nPVI and familiarity rating excluding the effect of the syncopation level, dropped (syncopation level $r = -.81$, score nPVI $r = -.53$), suggesting that both indexes explain parts of the variance independently.

Discussion

The result show that the participant groups were properly matched in terms of their exposure to the rhythmic patterns in the stimuli, because no significant difference on the familiarity ratings of the scores between the cultural groups was found.

The high correlation between score count and the familiarity rating provided an external validation of this measure; the familiarity measure may indeed reflect the level of exposure to the various rhythms. Therefore, it would be interesting to examine the performances of people who are exposed exclusively to one kind of music, because they might develop an idiosyncratic familiarity with rhythm, and as a consequence, behave in a different way.

The familiarity rating was highly correlated with the perception and production data, which might suggest that familiarity plays a part in the process. As in psycholinguistic studies, in which experimental control of, e.g., word frequency, is commonplace, perception and production studies on musical rhythm could also benefit from a tight control over this aspect.

Relating Perception, Production and Familiarity

Method

A Bayesian approach was applied to relate perception and production data. Central in Bayesian modeling is the notion of conditional probability, written as $p(a\,|\,b)$, which gives the probability of a occurring when it is given that b occurs. Bayes' rule relates the probabilities $p(a\,|\,b)$, $p(b\,|\,a)$, $p(a)$ and $p(b)$ as

$$p(a|b) = \frac{p(b|a) \times p(a)}{p(b)} \qquad (3).$$

Taking a to stand for rhythm categories and b for performances. We already applied the formula to relate rhythm perception and production of two-interval rhythm (see Sadakata et al.(2006). With this approach, $p(a\,|\,b)$, the probability of a rhythm (category) being perceived, given a (presented) performance, is predicted from $p(b\,|\,a) \times p(a)$, the probability of that performance being produced given a rhythm category as instruction, multiplied the prior probability of that rhythm category, divided by $p(b)$, the probability of the performance arising in any case. In the current study, by taking the familiarity rating as a candidate prior probability $p(a)$, we can predict the perception data from the production data distributions of three interval rhythms.

Materials

The production data distributions served as input, $p(b\,|\,a)$. They were calculated using Parzen method with a θ =0.04. The averaged familiarity rating for each culture was used as a prior probability. A uniform prior probability, $p(a)$ = 1/38, which does not differentiate between the rhythmic categories, was tried for comparison.

Results

The top four triangular plots in Figure 5.12 show the Bayesian predictions of the perception data of all categories, constructed with a uniform prior and a familiarity prior for each cultural group. The bottom two plots show the observed perception data of Experiment 1. As in the earlier perception plot, the gray surfaces correspond to the areas where the single major category is identified, and white surfaces to the areas where the majority response proportion was not significantly above the second most common response.

The agreement of the tiling of the categories, expressed as a percentage of performance space, increased to 61% for the Dutch group and to 69% for the Japanese group. The overlap in the tiling of a Bayesian prediction with uniform priors and the actual perception data corresponded for more than 50% (53 % for Dutch and 54% for Japanese). The crucial difference between the prediction and the observed perception data was that no large area of 1-1-1 category was predicted with uniform priors. By applying the familiarity ratings as priors, this major difference disappeared: The predicted size of the 1-1-1 category was in accordance with the observed perception data.

Discussion

In perception, the unequal competition between patterns may have been a large effect, obscuring the more subtle differences we find in production responses. Bayes' rule neatly explains how perception and production data may be different while the underlying distributions are still the same. "Simple" and "stable" rhythm categories seem to win as the preferred representation for a perceived performance when competing with other, more complex rhythms in perception (Desain & Honing, 2003), thus they cover a relatively large area in the triangular plot. However, in production, these patterns tend to be produced precisely and cover a small area on the same plot. This difference in the data characteristics tends to be overlooked when statistics, such as means and standard deviations, are compared. However, the Bayesian approach seems to be able to account for this difference.

The difference in data characteristics between perception and production is especially true for the isochronous patterns, 1-1-1, which usually exhibits one of the widest perceptual distributions, and the narrowest one in production. In the

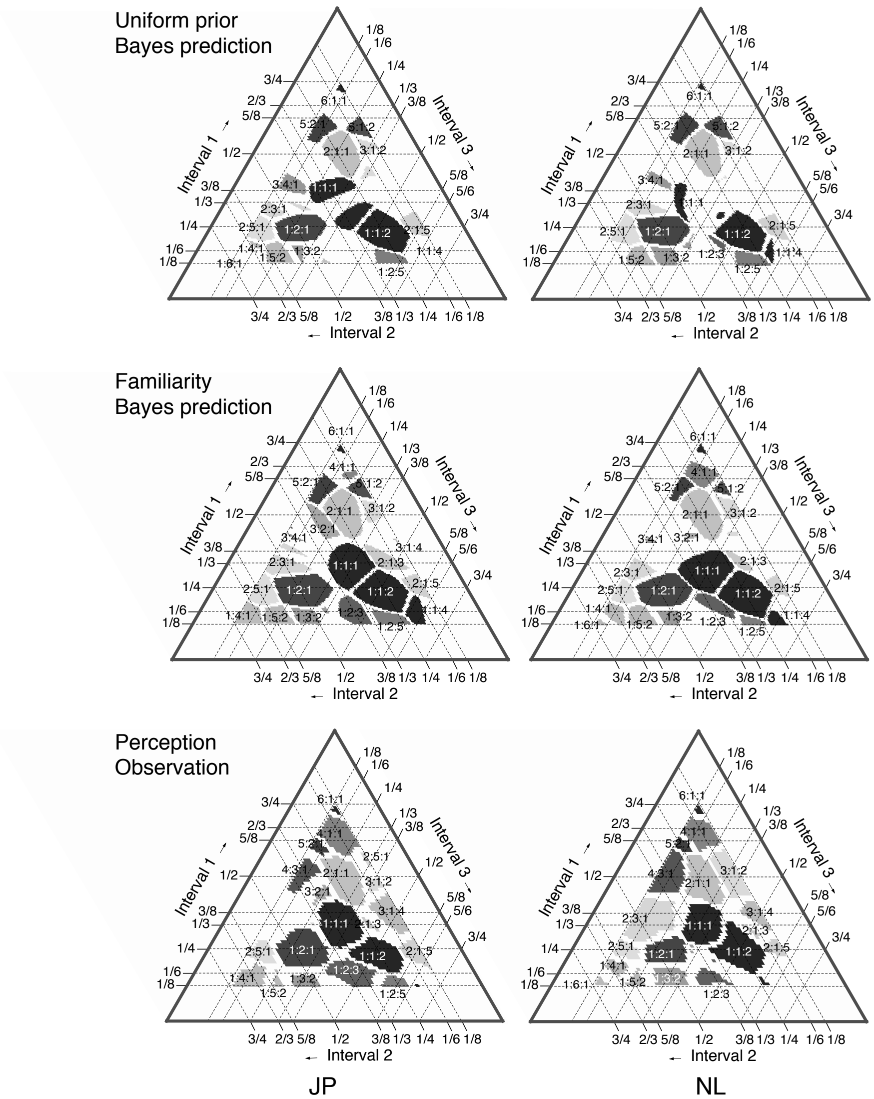


Figure 5.12 Observed perception results for the two cultural groups and the Bayesian predictions. As in Figure 5.7, the gray surfaces correspond to the area where the proportions of the significant major responses of the labeled patterns were identified, and white area to the area where a majority was not significant. The plots represent the pooled results for all time signatures. The top plots present the Bayesian prediction using uniform priors, the following row presents the prediction with familiarity priors, and the bottom row presents the observed perception result from Experiment 1.

current study, the Bayesian prediction with uniform priors failed to predict this middle area well. However, by assigning a large prior weight stemming from the familiarity rating as prior probability, the formula predicts the central area of the plot correctly. Not only predicting the size of majority areas, the Bayes rule also predict shifts in position caused by asymmetric strength of neighbor patterns.

Thus, in accordance with the previous study for two interval data (Sadakata et al., 2006), the Bayesian approach translates the difference in distribution characteristics between perception and production properly for three interval data as well, using independently collected priors from familiarity data.

General discussion

Throughout the experiments, higher correlations were observed between data and syncopation level, a complexity measure based on hierarchical structure, than between data and the score nPVIs, a serial complexity measure. Syncopation level also accounted for other aspects of the data, such as consistency across experimental results and ratings. Furthermore, the two cultural groups produced some of the three-interval scores with different timing profiles and this difference in timing became more prominent as the degree of syncopation increased, but was unrelated to nPVI. We take these results as evidence that rhythm is processed hierarchically rather than serially. In fact, the theoretical syncopation level accounted for data variability as much as did the familiarity ratings. Thus, mental representations based on a hierarchical structure might be formed as a result of exposure to the music that is organized in hierarchical way. If there is such an association between familiarity and mental representation, cognitive complexity may be non-universal. Thus, we should carefully state that this evidence of hierarchical rhythm processing was confirmed for a Western music context.

One issue that still needs to be addressed concerns the reason for the finding that culturally specific characteristics were only distinguished in the production of rhythm and not in its perception. It might be the case, as Sternberg, Knoll and Zukofsky (1982) proposed, that the perception and production processes do not share cognitive paths. From this perspective it is not surprising to find inconsistent results for perception and production. However, many authors have pointed out that such a view is not realistic

(Drake, 1993; Penel & Drake, 1998; Repp, 1995; Sadakata et al., 2006). Rather, perception and action are commonly thought to be closely related (Hommel, Musseler, Aschersleben, & Prinz, 2001; Rizzolatti, Fadiga, Gallese, & Fogassi, 1996). According to these views, the discrepancy between perception and production is puzzling. Although we cannot provide an appropriate explanation for the absence of a cultural difference in perception while the effect was present in production, we still take the standpoint that perception and action are closely related, and furthermore, we propose that they are related in a Bayesian way.

In the current study, instead of the traditional way of comparing mean interval durations of perception and production, the Bayesian approach to relate perception and production proposed by Sadakata et al. (2006) was used and shown to function well. Considering the observation that the within-subject consistency of the perceptual experiment was around 70%, we think that the agreement between Bayes prediction with familiarity priors and the observed perception data is quite high (65% on average). Thus, if perception and production are indeed related in a Bayesian way, the result confirmed that the rhythm categories have non-uniform prior probabilities, and, familiarity rating is shown to be a good independent measure of prior probability. Together with the close relation between familiarity rating and the score counts, it suggests that these priors are indeed learned from the environment and the experience of the individuals.

Finally, there is the question of the origin of the observed culturally specific tendency found in rhythm production. The cultural diferences we found depended on the position of the intervals, which is different from the correlation between music and speech in terms of their durational relation of successive interval (Huron & Ollen, 2003; Patel & Daniele, 2003b; Sadakata & Desain, submitted-a). Indeed, this positional information in temporal sequences has been shown to play a part in producing temporal events. For instance, it is well known that vowel durations as well as musical note durations are lengthened at the ends of phrases (Beckman, Edwards, and Fletcher, 1992; Honing, 2003; Sundberg and Verrillo, 1980; Todd, 1985). More interestingly, Kubozono (2002) reported a culturally specific tendency in the attention to positional information. In his study, Japanese speakers were better than English speakers in discriminating the length of the vowels that appeared at the beginning or in the middle of an utterance, while they performed poorly in discriminating the length of the end vowels. Therefore, the position-oriented culture-specific tendency in music rhythm production found in the present study might arise if a different sensitivity in positional information is required for processing different languages.

Conclusion

We found a meaningful relation between a measure of cognitive complexity and the perception and production of rhythmic patterns. Most importantly, we demonstrated that, at least within a Western musical context, the cognitive processing of three-interval rhythms builds on a hierarchical mental structure, as shown by the positive correlation between a measure of the complexity of the rhythm and the amount of diversity in processing musical rhythm. The variability within and between individuals became larger when the task involved more complex rhythms. Furthermore, the observed variability became culturally specific in some cases. Dutch and Japanese participants showed different timing profiles in the production of rhythms that contain two syncopations. Furthermore, we demonstrated that, incorporating familiarity, the different characteristics of the data of perception and production could be successfully translated and associated by a Bayesian formula.

Chapter 6

Learning to perform simple rhythmic patterns with expressive deviations

Abstract

Learning to perform music expressively involves complex processes, such as acquiring a sense of style and sensitivity to expressive variations made by other performers. To our knowledge, there have been no studies done on how quickly expressive music performance can be learned. We studied learning as well as transfer of learning related to this issue. Twelve amateur musicians participated in the experiment, which consisted of a pre-test, two training sessions and a post-test. In the training sessions, participants trained the imitation of short rhythms with expressive variations. In the pre and post-test, participants performed an imitation as well as a perceptual discrimination task. Improvement of expressive performance was measured by comparing performance in the pre and post-test. Analyses of the imitation task revealed an improvement of task performance not only for the trained renditions but also for non-trained renditions. The degree of improvement varied with stimuli type and dimension of the responses (timing or loudness). An improvement of perceptual accuracy was not observed. The results provided insight into the issue of which aspects of expressive music performance share a common basis; certain timing variations are closely associated and show larger transfer of learning. Transfer of learning does not behave in the same way for timing and for loudness, which suggests that these parameters are not likely to be integrated into a single dimension of musical expression.

Artistic expression is one of the crucial factors that make music enjoyable. Musical expression has been studied by looking at, for example, deviations in the timing and loudness of a musical performance from the information specified in the musical score. Various authors have shown that expressive deviations are systematically linked to diverse aspects of the music performance, such as structures (Gabrielsson, 1974, 1987; Sundberg & Verrillo, 1980; Todd, 1985; Windsor & Clarke, 1997), performance styles (Collier & Collier, 2002; Friberg & Sundström, 2002; Repp, 1990), performer's intentions (Kendall & Carterette, 1990; Palmer, 1989), and performer's cultural backgrounds (Ohgushi, 2002; Sadakata, Ohgushi, & Desain, 2004). Concerning processes underlying the learning of the performance music, several studies have investigated the impact of musical training on the processing of musical information (Drake, 1993a, 1993c; Drake & Botte, 1993; Drake & Palmer, 2000; Palmer & Drake, 1997). However, the present study focuses on how fast musically trained individuals are able to learn varying ways of performing music and to what extent do learned abilities transfer to novel material and novel tasks.

In this study, the term 'transfer of learning' refers to a change in response to a novel material and novel task as a result of a previously experienced task. This process can be seen as transfer of acquired skills from one situation to another. The success of transfer of learning can be regarded as an indication of an ability to abstract and to generalize the skills of the learned tasks (Meyer & Palmer, 2003). This approach is often applied to study performances of different tasks, for instance, solving algebraic problems (Robertson, 2000), applying learning strategies (Scruggs & Mastropieri, 1988), as well as the learning of grammatical rules (Marcus, Vijayan, Rao, & Vishton, 1999), phonological categories (Maye, Werker, & Gerken, 2002), event sequences (Cohen, Ivry, & Keele, 1990), and control of motor patterns (Heuer & Schmidt, 1988), to name a few.

As music is a complex domain that requires wide-ranging motor-skills as well as cognitive operations, learning how to perform music certainly requires learners to transfer previously learned skills to novel situations and materials in various ways. For example, Palmer and Meyer (2000) studied how adult musicians and children transfer learned skills of piano performance to new situations. The two new conditions differed from the practice condition either in conceptual (pitch sequence) or in motoric (hand and finger) aspects. An interaction of the conditions and the musical experience was found on the success of transfer: adult musicians transferred their learned skills best to the condition that retained conceptual similarities rather than the one that retained motoric similarities, and the opposite was true for the least experienced children. Thus, they showed that the conceptual aspects of music performance become independent from the required movements at advanced skill levels. Using the same paradigm with piano performances of simple melodies, Meyer and Palmer (2003) found a different amount of transfer between the conditions under which motor aspects (hand and finger) or temporal aspects (rhythm and meter) were altered. This finding implies independent motor and temporal representations of piano performances. Thus, a transfer of learning paradigm is a powerful way to tease apart various aspects of the cognitive processes underlying music performance.

The main focus of the current study is on the learning process of the skills needed to perform expressive musical renditions. Apparently, learning of musical expression occurs at different stages. For instance, a slow learning process takes place as a result of exposure to environment information. More experience with certain types of music enhances one's knowledge of that particular music. Therefore, listeners often increase their sensitivity to the structural information contained in a more familiar style of music (Drake & El Heni, 2003; Hannon & Trehub, 2005). Furthermore, it has been suggested that exposure to non-music materials such as speech may play a role in perception of musical materials (Iversen, Patel, & Ohgushi, 2004; Stevens, 2004; Stobart & Cross, 2000). Not only perception, but also experiences of different environments seem to have an impact on the way musicians produce simple rhythmic patterns (Sadakata & Desain, submitted-b; Sadakata et al., 2004). On the other hand, there is a faster and more active kind of learning that takes place during music lessons and practice sessions. Little is known about how quickly expressive renditions are learned and transferred in these kinds of situations. Therefore, our first question concerns whether the skills needed to perform expressive renditions can be learned after a short period of training, and whether the acquired skills transfer to the non-trained renditions.

To be able to recognize two superficially different rhythms as different expressive renditions of the same score, one needs to know the general structure of the rhythm regardless of the surface structure that is unique for specific expressive renditions. "General structure" here refers to score information represented in integer durational ratios, which are hierarchically structured into subdivisions in western tonal music. If these common mental representations are not activated, the two renditions are recognized as two different rhythms. It has been shown that general structures of rhythms form categories that can be abstracted from their specific timing profiles (Clarke, 1987; Desain & Honing, 2003; Schulze, 1989). This explains why rhythms that have different surface structures are considered to belong to the same category. Similar examples can also be found in music production. For example, professional musicians can perform the same scores with different expressions (Windsor & Clarke, 1997), suggesting that the same rhythm category is applied to construct a different surface structure. These examples suggest that the various renditions that share common rhythm representations are likely to be more closely connected as mental representations. This leads to our second question, whether the skill of performing rhythms transfers more easily to the different rhythms belonging to the same categorical representation than to rhythms belonging to different categories.

The third question concerns whether the skills used to produce musical rhythms also transfer to a perception task. It is plausible that transfer is facilitated when the conditions of the practice and the transfer tasks are more similar. Indeed, previous work shows that the more the tasks differ, the smaller the transfer of learning (2000). Transfer between different tasks indicates that the abstraction of knowledge or skills that are independent of the particular task takes place through training. Motor skills involved in production and perception tasks differ considerably, and, furthermore, the nature of the tasks is different (Sadakata, Desain, & Honing, 2006). Therefore, a transfer effect from production to perception suggests not only the independence of transferred skills from the motoric aspects of the production task, but also may suggest the existence of a common process involved in the tasks. For example, the skill to perceive rhythm is quite likely to be involved in the imitation task. Such transfer from production (imitation) to perception has been reported in speech research (Hirata, 2004).

The issues raised here are important for at least two reasons. First, it could add considerably to our understanding of musical performances. Investigating the transfer of performance skills across tasks and stimuli could provide important information about which aspects of expressive music performance involve the common process. Furthermore, the study is helpful in the development of new computer training programs for musical performance, because the knowledge of the transfer process of the skills needed to perform expressively may suggest an efficient way to practice diverse musical expressions.

For the current study, a transfer of learning effect paradigm is employed to assess improvement in the production and perception of musical performances with expressive variations. The speeding method, as used in previous studies (Meyer & Palmer, 2003; Palmer & Meyer, 2000), is not optimal for the purposes here, because speeding severely restricts the possible range of expressive renditions. Instead, an imitation task was utilized, in which accuracy of imitation, measured as the inverse distance between target and response, served as an index of performance skills.

Among the many possible musical materials, we opted for short musical rhythms, more specifically three interval rhythms with four note onsets, because of the accumulated knowledge regarding the processing of these rhythms (Desain & Honing, 2003; Sadakata & Desain, submitted-b). We used expressive variations in two domains, timing and loudness, as they are commonly used as parameters of musical expressions (Repp, 1998, 1999a; Timmers, 2005; Windsor & Clarke, 1997). A pre-test post-test within-subject design was utilized. The pre- and the post-test were identical, consisting of an imitation and a discrimination task. The accuracy of a one-shot imitation was measured in the imitation task, while a same/different judgment was used in the discrimination task. In between the pre- and the post-test a training session took place, which contained only the imitation task. The test stimuli involved musical rhythms with expressive variations, created by augmenting the performance scores with artificial expressive deviations derived from a previous study on the production of similar three interval rhythms (Sadakata & Desain, submitted-b): performances with averaged timing and loudness deviations (prototype), or performances with either timing or loudness variations. A subset of the test stimuli was used during the training session. In the tests, two types of transfer conditions were included. The first, referred to as the non-trained-score stimuli, presented a rendition of a musical score that was not trained for at all. The second stimulus type presented a rendition with an expressive pattern that was not trained for but a score that was trained with another expressions, which we refer to as the non-trained-expression stimuli.

The rms (root mean square) error between the target performance and the responses was calculated for timing and loudness separately. An improvement of the imitation task from pre-test to post-test for trained renditions indicates a learning effect, while an improvement for non-trained patterns indicates a transfer of learning effect. The main interest was to see whether the learning effect as well as the transfer of learning effect would take place. We expected to find a larger transfer effect for non-trained-expression stimuli than non-trained-score stimuli, as it was expected that practiced skills would be more easily transferred within the same rhythm category. An increased sensitivity in discrimination as a result of the imitation task is expected too, because perception is involved in the imitation task and may be implicitly trained as a result of the imitation training.

Method

Participants

Twelve paid Dutch participants took part in the experiment. All of the participants were amateur musicians who had received on average 13.3 years of formal musical training (ranging from 10 to 16 years). Their mean age was 25.9 years (ranging from 17 to 42 years). Amateur musicians were preferred over non-musicians as we are interested in the learning process during the musical lessons or practice sessions.

Material

Two-bar rhythms were used as stimuli for both the imitation and the discrimination tasks. The first bar contains three intervals (notes) and the second bar contains one note. The bar duration was one second, thus each pattern lasts two seconds. The onset of the first note in each bar was marked by a metronome sound. The categorical durational ratios of the three notes (interonset intervals, IOI) in the first bar were permutations of 1-2-3, which resulted in six patterns. All rhythmic patterns were presented in a time signature of 3/4 and 6/8, counting for a total of 12 scores. Sadakata & Desain (submitted-b) collected production and difficulty rating for the same rhythms from 36 conservatory piano students. The 12 scores and their difficulty ratings are shown in Table 6.1.

Figure 6.1 shows nine types of expressive renditions that were applied for the 12 scores. The nine types of variations were acquired by systematic variation of two parameters. These two parameters of performance, namely inter-onset-interval (IOI) and loudness, were modified to construct diverse renditions. The range of the variations was taken from the rhythm production data described by Sadakata & Desain (submitted-b). The prototype rendition was defined as the average performance from that study: performance of which IOI and loudness corresponds with observed means indicated as timing = = =, loudness = = = =. Patterns with a timing accent and a loudness accent were created. The timing accent was created by shifting one onset later (+ - =, = + -) or earlier (- + =, = - +), while the loudness accent was made by making one note louder (+ = = =, = + = =, = = + =, and = = = +). The plus sign in the timing accent indicates that the corresponding IOI was prolonged, while the minus sign indicates it was shortened. For the loudness, the plus sign indicates that corresponding interval was louder than the other. Only one accent in one parameter was considered at a time.

Table 6.1 Twelve scores and their difficulty rating used for both imitation and discrimination tasks. The difficulty ratings were derived from Sadakata & Desain (submitted-b).

Meter	Pattern	Difficulty
3/4	1-2-3	0.62
3/4	1-3-2	0.36
3/4	2-1-3	0.41
3/4	2-3-1	0.15
3/4	3-1-2	0.17
3/4	3-2-1	0.62
6/8	1-2-3	0.45
6/8	1-3-2	0.61
6/8	2-1-3	0.28
6/8	2-3-1	0.4
6/8	3-1-2	0.46
6/8	3-2-1	0.25

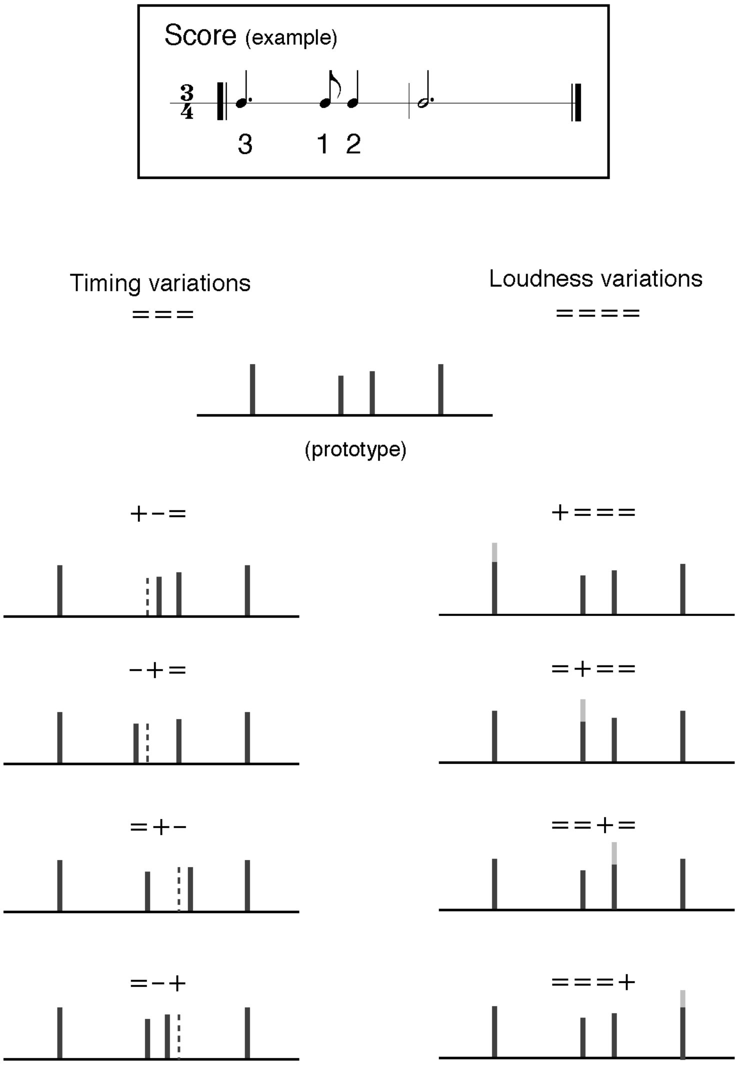


Figure 6.1 Nine types of expressive renditions. The equal signs for both timing and loudness accent columns indicate the means derived from Sadakata & Desain (submitted-b). The plus sign in the timing accent column indicates that the corresponding interonset intervals were prolonged, while the minus sign indicates shortening. The plus sign in the loudness accent column indicates that loudness of the corresponding notes was increased.

The modulation range of the accents should be large enough so that different renditions are clearly distinguishable, but it should not be too much to make sure that renditions are still associated with the same score. For the imitation task, this range was taken to be proportional to the data distribution of production timing, which were calculated using the Parzen method (Parzen, 1962). Figure 6.2 illustrates the distribution of the data that encloses 90 % of the performances (3-1-2 pattern) as a gray ellipse in the performance space, which contains all possible performances of four onsets with a total duration of one second (see Desain & Honing, 2003 for a detail explanation of the plot). The black dots at the boundary of the gray area in the figure indicate the selected renditions with expressive timing accent. For the loudness accent, the velocities were 3.5 standard deviations (SD) louder than the mean level (approximately 2-3 dB).

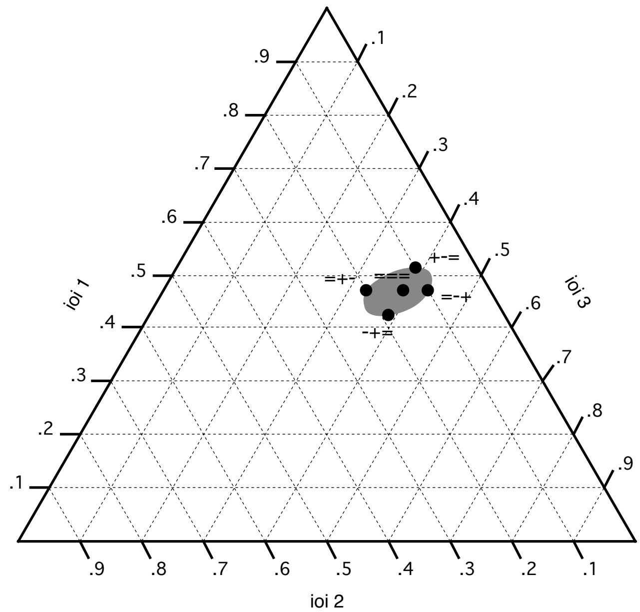


Figure 6.2 Triangular performance space which express all possible performances of four onsets with a total duration of one seconds (see Desain & Honing, 2003 for a detail explanation). An example distribution of the data that encloses the 90 % of performance is shown as gray circle. The renditions on the boundary of the distribution, shown as black dots were used as the expressive timing accent.

Imitation task (Pre-test/Post-test, training)

Time signature, rhythmic patterns, the difficulty rating, and the patterns of the expressive renditions were counter-balanced across participants and were assigned as follows. Table 6.2 presents the overview of the stimuli sets for training and for tests. The training subset consisted of three expressive renditions, one with a prototype profile, one with a timing accent and the other with a loudness accent for each score. As four scores were assigned for the training session, the set contains 12 stimuli: three renditions for four scores. The test subset consisted of the training subset plus three non-trained stimuli types: two non-trained expressive renditions, one with a timing accent and one with a loudness accent, and one non-trained score rendition, the prototype rendition of four new (untrained) scores. Thus, the set contains 24 stimuli: five types of renditions for four trained scores and the prototype renditions of four untrained scores.

Table 6.2 Overview of the stimuli used and the stimulus type for the training and tests. Prototype refers to the rendition of which IOIs and loudness were average values, while timing variation, loudness variation refers to the renditions of which timing or loudness were altered. The expressive ranges were obtained from the study by Sadakata & Desain (submitted-b).

Condition	Number of scores	Stimuli type		
training (12 stimuli)	4	trained		prototype
				timing variation
				loudness variation
test (24 stimuli)	4	trained		prototype
				timing variation
				loudness variation
		non-trained	expressions	timing variation
				loudness variation
			scores	prototype

Discrimination task (Pre-test/Post-test)

A different modulation range was applied for the discrimination task, as the accents should not be too obvious in this case. To construct stimuli, 24 ms onset shifts as timing accent and 1.42 SD loudness level increase as loudness accent were used. This difference in loudness roughly corresponds to 1-2 dB in sound pleasure level. These values were at 50% correct level of the discrimination experiment conducted separately with five amateur musicians who had on average 12 years of musical training.

The stimuli consisted of two conditions: the 'different' condition, in which the renditions between the pair of the rhythms presented were different, and the 'same' condition, in which the renditions were the identical. Rhythms with three types of expressive renditions were included in the trainings: one with a prototype profile, one with a timing accent, and the other with a loudness accent. The renditions that included a timing or loudness accent were always paired with prototypes to constitute the 'different' condition, while the 'same' condition consisted of a pair of prototypes. As four scores were assigned for the test, the set contained 16 stimuli: two 'different' conditions and two 'same' conditions for each of the four scores.

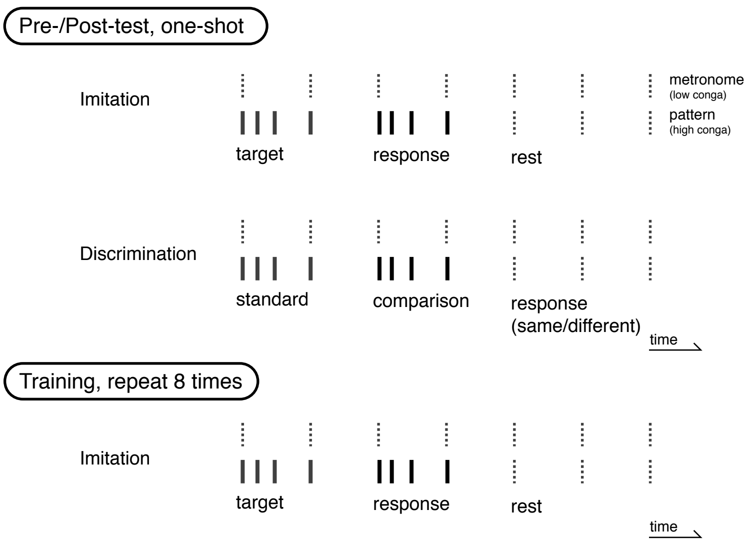


Figure 6.3 An illustration of the tasks for tests and trainings.

The position of the two pads could be changed: the red pad was always set for the preferred hand. Tapping the green pad with the index finger allowed instructions and stimuli to be presented for both the imitation and discrimination tasks. Figure 6.3 illustrates the tasks. In the imitation task, the target performances to be imitated were presented during the first two bars, and the subsequent two bars were provided for participants to give a response with the red pad, followed by another two empty bars to provide a rest. In the discrimination task, the first rhythm pattern was presented in the first two bars, followed by the second rhythm pattern in the subsequent two bars. Participants gave responses during the last two bars by hitting the green (same) or the red (different) pad. The responses of the participants were presented in the display. They could change their responses as often as they wanted during the response period (2 seconds). Scores of the target pattern were presented throughout the experiment.

Participants were given the opportunity to practice the task 3 times in order to get used to the procedure, using different rhythms (in 4/4 time signature). The task was self-paced. Including breaks, the pre- and post-test took participants from approximately 10 minutes to complete and each training session also took about 10 minutes.

In the pre-tests, the imitation task was presented first, followed by the discrimination task. The training session (imitation task) was repeated, using the same stimulus set presented in the same order. The training sessions were followed by the post-test. Pre- and post-tests were identical, with the exception that the practice sessions were removed from the post-test.

Results

Imitation task

Pre-processing

The MIDI files were converted to tables of IOIs of the performance using POCO. We discarded trials with more or fewer than four onsets and trials that had a total duration outside the range of 500 to 1500 ms. Seven of 288 trials did not fulfill the criteria and were discarded for the pre-tests, one of 288 trials for the post-tests, while 30 and 23 of 1152 trials for the training sessions 1 and 2, respectively. The rms errors for timing and loudness were computed and submitted for analysis. The timing error was calculated for the four onsets of the targets and the responses. Similarly, the loudness error was calculated for the four notes of the targets and the responses. Outliers, which were outside three standard deviations from the means, were computed across participants for the pre- and post-tests, and for two training conditions. Six of 288 trials for the pre-tests, six out of 288 for the post-tests, 26 of 1154 for the first training sessions and 27 of 1154 for the second training sessions were discarded according to this criterion.

General

We tested whether learning and transfer of learning took place. The mean errors and standard errors of timing and loudness from the pre-test, and those for trained and for non-trained renditions from the post-tests were plotted in Figure 6.4. The degree of the transfer of learning was measured as the difference between the rms errors of the pre- and post-tests. Test stimuli were classified into the following four types: 1) *trained*, rhythms presented in the training session, 2) *non-trained-timing*, the trained scores combined with a non-trained timing variation, 3) *non-trained-loudness*, the trained scores combined with a non-trained loudness variation, and 4) *non-trained-scores*. A repeated measures ANOVA on both the timing and loudness rms errors by test (2 levels) and stimulus type (4 levels) revealed the significant main effect of test condition, for timing, $F(1, 265) = 16.18$, $p < .01$, and for loudness, $F(1, 265) = 12.94$, $p < .01$, with no significant interaction between test and stimulus type. There was no significant effect of the variation type, for timing, $F(3, 265) = 0.60$, $p = .61$, nor for loudness, $F(3, 265) = 1.28$, $p = .28$. This indicates that learning took place, not only for the trained renditions but also for the non-trained ones.

The error may decrease, as the participants get further into the experiment due to participants' ability to adjust their performance to the setup and tasks. However, there were no significant correlations between the errors and the number of trials or the number of blocks for both trained and non-trained conditions. This suggests that the improvements of the imitation task found in the post-tests for both trained and non-trained conditions were not due to the participants' general adaptation to the task.

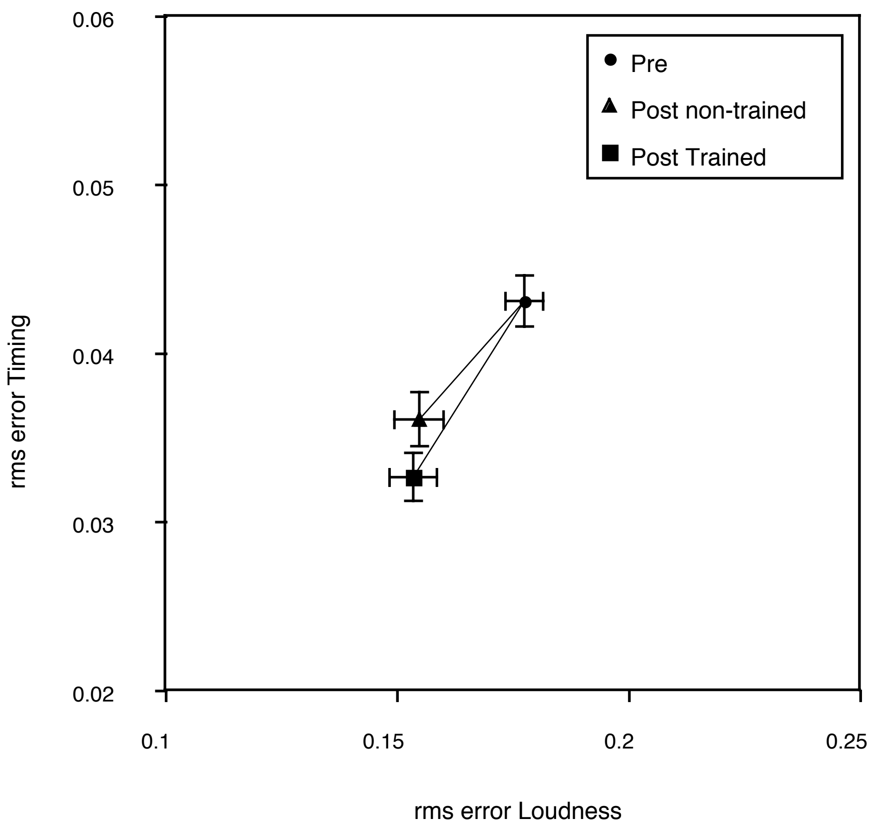


Figure 6.4 The means and their standard errors of the timing and loudness rms errors for pre-test and for post-tests. All stimuli types are presented together for pre-test condition as circle, while trained and non-trained renditions are separately shown for post-test condition.

Timing

Detailed analysis of the timing rms errors was conducted on the data excluding the loudness variation renditions. First, the pattern of learning during the training session was studied. This type of study often reports a decrease in the error measures as the number of the trials increases. Slight but significant negative correlations between the number of training trials and the rms error were obtained (Spearman's Rho; Training 1, r = -.08, $p < .05$; Training 2, r = -.12, $p < .01$), suggesting that the learning took place as the number of the trained trials increased. The two analyses suggest that improvement was greater in the second training session than the first training session.

Figure 6.5 shows the mean timing rms errors and standard errors of pre- and post-tests, as well as the first and the last trials of the two training sessions, for different stimulus type. The errors of the renditions that were presented in the training session are labeled as *trained stimuli*. The other renditions were classified as *non-trained stimuli*. For the non-trained stimuli, a distinction is made between *non-trained-timing*, which includes performances containing non-trained timing variations on previously trained scores, and *non-trained-scores*, which includes the prototype renditions of the non-trained scores. The errors of all stimulus type are shown together for pre-test, because there was no distinction between stimulus types before the training.

A t-test on the errors comparing the renditions with accent and without accent of the pre-test indicated no significant difference, t(180) = 1.6 p =.11. This suggests that to imitate accented renditions was not more difficult or easier than imitating prototype renditions in the pre-test. Therefore, we consider that the modulation of the timing accent was in a natural range.

Although the presence of the transfer effect for all the stimuli has already been confirmed in the general analysis part, more detailed examination of the effect sizes within non-trained conditions is of interest here. Two paired t-tests that directly contrast the results of pre- and post-tests for non-trained-timing stimuli and for non-trained-score stimuli were conducted for this purpose. Obviously, both data sets of stimuli showed a general tendency of the errors to be smaller for post-test than for pre-test. For the non-trained-timing stimuli, the difference was of borderline significance, t(47) = 1.63, p = .055, while effect was smaller for the non-trained-score stimuli, t(45) = 1.12, p = .13. Accordingly, the comparison of the significant levels between two tests suggests that skill of imitating rhythm transfers more easily to the rhythms that have the same categorical representation than to the rhythms that belong to different categories.

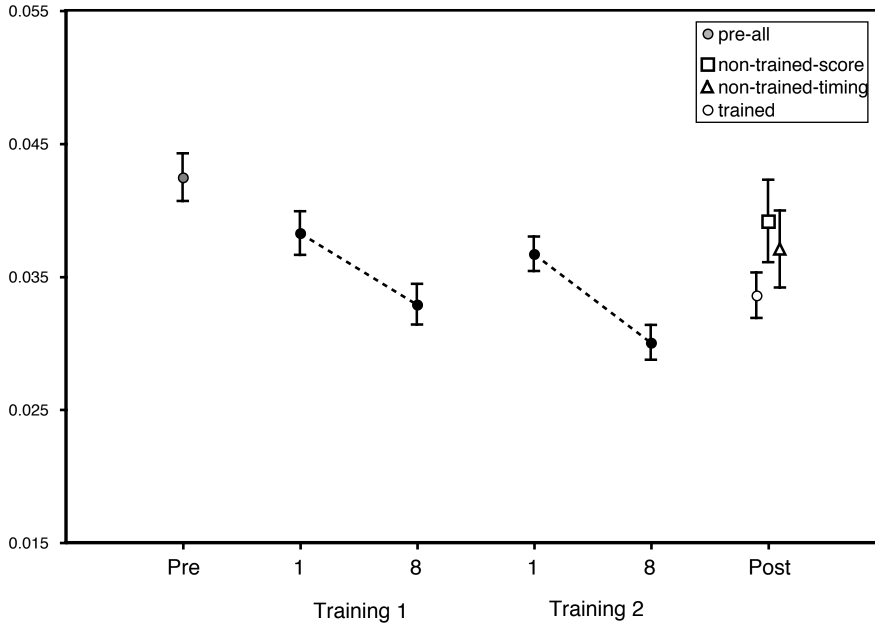


Figure 6.5 The mean timing (rms) errors and standard errors of pre- and post-test, and the first and the last repetition of two training session for different conditions. The errors of the renditions that were presented in the training session are labeled as trained. Within the non-trained conditions, the errors for the trained scores combined with a non-trained timing variation were shown as non-trained-timing, while that for the prototype renditions of the non-trained scores were shown as non-trained-score.

Loudness

Detailed analysis of the loudness (rms) errors was conducted on the data excluding the timing variation renditions. The pattern of learning was first studied. The number of training trials and the errors indicated no significant correlations (Spearman's Rho; training 1, $r = -.00$, $p = .93$; training 2, $r = -.03$, $p = .49$). The results suggest that the learning curve was not simply a function of the number of training trials.

Figure 6.6 shows the mean loudness rms errors and standard errors of pre- and post-tests, and the first and last trials of the two training sessions, for different stimulus types in the same manner as Figure 6.5. Within the non-trained stimuli, distinction is made between non-trained-loudness stimuli, the trained scores, and non-trained-score stimuli, the prototype renditions of non-trained scores.

A t-test on the errors comparing the renditions with accent and without accent of the pre-test indicated no significant difference, $t(181) = 1.343$, $p = .18$. This suggests that to imitate accented renditions was not more difficult or easier than imitating prototype renditions in the pre-test. Therefore, we consider that the modulation of the loudness accent was in a natural range.

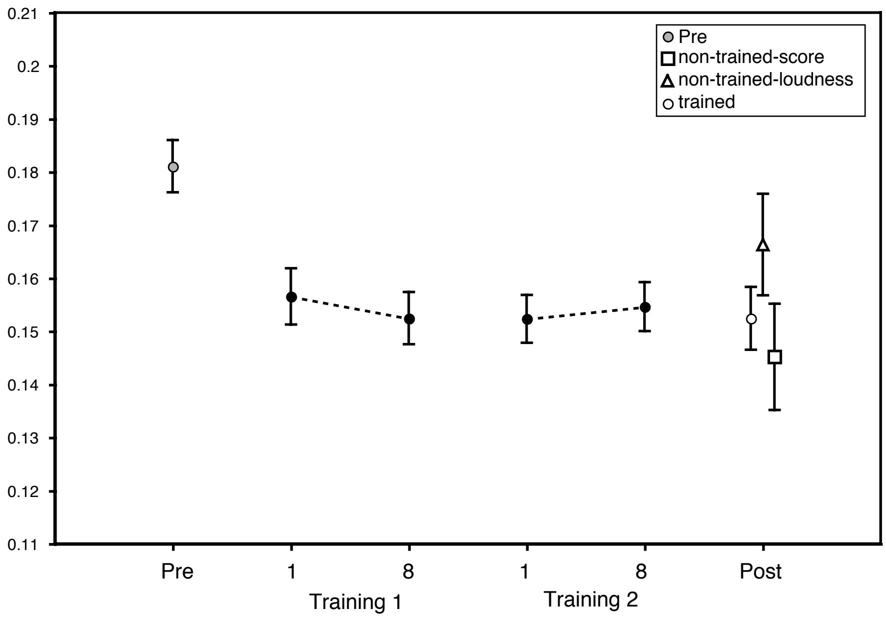


Figure 6.6 The mean loudness (rms) errors standard errors of pre- and post-test,and the first and the last repetition of two training session for different conditions. Theerrors of the renditions that were presented in the training session are labeled as trined. Within the non-trained conditions, the errors for the trained scores combined with a non-trained loudness variation were shown as non-trained-loudness, while that fr the prototype renditions of the non-trained scores were shown as non-trained-score.

As shown in Figure 6.6, the first trial of the first training session already marked a significantly smaller error than the two pre-test conditions, t(269) = 3.141, $p < .01$. It is difficult to explain the reason why the errors for non-trained-scores stimuli were small.

As in timing, we did a more detailed examination of the effect sizes for non-trained conditions. Two paired t-tests that directly contrast the results of pre- and post-tests for non-trained-loudness stimuli and for non-trained-score stimuli were compared. Both data showed a general tendency of the errors to be smaller for post-test than for pre-test. For non-trained-loudness stimuli, the effect size was rather small, t(43) = 1.33, $p = .09$, while a more significant effect was confirmed for non-trained-score stimuli, t(45) = 1.79, $p < .05$. Therefore, unlike the timing responses, the comparison of the significance levels suggests that the skill to imitate loudness rendition transferred more easily to the rhythms belonging to different categories than to the rhythms that belong to the same categories.

Discrimination task

The proportion of correct responses for the discrimination task was 77.1 % for the pre-tests, and 76 % for post-tests. To study the differences in sensitivity to renditions of different timing and loudness variations in the pre- and post-tests, a D-prime measure was calculated for each test condition for each participant. Because error rates for all tasks were very low, z-scores for false-alarm rates of 0 were approximated to be -4. If the patterns of score timing or loudness variation were learned through imitation training, this knowledge might be transferred to a different task. Higher D-primes for the post-test than for pre-test would suggest this transfer effect. A paired t-test revealed that D-primes were not significantly different between pre- and post-tests, t(11)=0.45, $p = .33$, indicating that participants did not improve their sensitivity after the training. Therefore, the results did not confirm the transfer of the skills between different tasks.

Discussion

In the current study, we have shown that the skill to imitate musical expression can be learned and improved through training rather quickly. Furthermore, transfer of learning was observed to the renditions of both trained and non-trained short rhythm patterns. An improvement of task performance for the non-trained trials was confirmed to various degrees for different stimulus types. Finally, no transfer was confirmed from the imitation to the discrimination task.

Some results, such as the smaller errors for the second training compared to the first training with regard to timing responses, may imply that the results did not indicate an improvement in skills to imitate renditions, but instead might indicate participants' ability to adjust their performance to the setup and tasks. However, for both tests and trainings, there was no significant correlation between the errors and the trial numbers or block numbers. Thus, getting used to the setup and tasks cannot be the whole story.

Regarding the timing responses, a larger transfer effect was observed for non-trained-timing stimuli than for non-trained-score stimuli. Thus, the skill to imitate expressive timing transferred more easily to rhythms that have a different surface structure but belong to the same categorical representation, than to rhythms that do not have a common categorical representation. In other words, to learn a way of timing a particular score was sufficient for musicians to improve their skill when imitating other expressive variations of the same score. This may indicate that different surface structures of rhythms characterized by a common categorical representation are closely connected and may even be processed in a similar way. If so, this may also suggest that the score and expression through timing are represented separately in the brain. This indirectly shows categorization of a similar surface structure of the rhythm into a score related mental representation.

However, it may also be the case that it was more difficult to apply learned skills to rhythms with greater differences than to more similar rhythms, regardless of the underlying categorical representations. Clearly, rhythms that belong to the same categorical representation are often more physically similar when compared to rhythms that differ in categorical representation. Thus, this issue still needs further consideration. Nevertheless, the results offer evidence for the transfer of the learning of performance skills at a more general level than just a surface structure specific to unique expressive renditions.

The pattern of the transfer effect observed for loudness responses contrasted with the one for timing responses. Regarding loudness responses, a larger transfer effect was observed for non-trained-score stimuli than for non-trained-loudness stimuli. Thus, the skill to imitate expressive loudness transferred better to the conditions presenting non-trained scores than to the ones presenting trained scores with non-trained loudness expressions. This different pattern of transfer between the dimensions of the musical expression may be due to the fact that the average loudness accentuation was more even. Therefore introducing loudness deviations changes the characteristics of the loudness profile, from a rather even to an uneven one. However, because the contrast of note durations was still maintained for averaged timing profiles, as for the permutations of durational ratio of 1-2-3, adding timing deviations to the average expression does not drastically change the characteristics of the timing profile, the uneven duration ratio stays uneven. Although this is speculative, such a difference in characteristics of changing profiles may have played a role.

Regarding the training sessions, different learning curves emerged for the timing and loudness responses. While the errors for timing decreased gradually during the training sessions, those for loudness already dropped at the first training trial. Even though this indicates a complicated issue, such non-parallels between expressive deviations of timing and loudness have been reported in other studies as well. For example, Windsor and Clarke (1997) showed that the expressive model proposed by Todd (1992) predicted the real musical performance best when different parameters for timing and loudness were assigned. Also, Timmers (2005) reported that models based on tempo or loudness do not equally account for how people judge the similarity of musical performances. Indeed, it is an intriguing question how the different aspects of musical expressions are represented and associated. The differences of the learning curves for timing and loudness observed in this study may suggest the independence of the learning process for these two components of musical expression.

The last issue that has to be mentioned concerns the lack of transfer from production to perception. The expected transfer effects were not confirmed in the study. This suggests that, although the learned skills were transferred to the non-trained imitation conditions, this transfer operation took place only at a task dependent level. This may be due to the short training time (about 20 minutes) used to train many similar conditions. An alternative explanation is a possible ceiling effect – a correct response rate of 77 % for the quite small deviations is already good, as compared to the expected 50 %.

The knowledge gained from music cognition research that investigates human perception and production is often directly applicable for pedagogical purposes. For example, the main finding of the current study, that the skills to perform expressive renditions can be learned and transferred rather quickly, is useful information when one designs computer assisted training programs. Also, the reported independence of the timing and loudness in the learning process is an important finding not only for teachers but also for when these parameters are presented as feedback in such a computer learning system. However, obviously, the present study is a first step with simple materials, and it raises many other questions as well. Many issues still have to be considered; for example, further exploration of a transfer effect from production to perception, a transfer effect from perception to production, and the relation between the amount of the training sessions and the size of the transfer of learning effect, etc. In fact, Travlos has demonstrated that the transfer effect decreases after a certain amount of practice. Therefore, more practice does not necessarily enhance the transfer effect (Travlos, 1999). It may be the case that learning becomes too specific after a certain amount of training. To answer these research questions is certainly fruitful for the domain.

Conclusion

The reported study showed that the skills involved in imitating expressive renditions of short musical rhythms can be quickly learned and improved. Furthermore, participants were able to apply learned skills to newly presented scores. The degree of the transfer depended on the type of stimuli and on the dimension of the responses (loudness or timing). The results provide us with insight into the issue of which aspects of expressive music performance share a common basis; first, certain timing variations are more closely associated than others, and second, timing and loudness parameters do not behave in a similar way and are therefore not likely to be integrated into a single dimension of musical expression.

Epilogue

The words *Ritme* [rItmə] and *Rizumu* [ɹizumu] mean 'rhythm' in Dutch and Japanese respectively. The words were chosen for the title of this thesis to allude to the distinct prosodic characteristics of different languages. In this thesis, we have investigated musical rhythm perception and production from various angles. One of the main aims was to examine the question of cross-domain influence, namely the effect of exposure to speech rhythm on music rhythm processing. We have found some cases where such an effect is apparent. Furthermore, the strength of the effect has been shown to interact with the complexity of the rhythmic structures. A more salient effect seems to emerge for rhythmic patterns that have a more complex structure and therefore require more complex mental coding. In addition to this cross-domain influence, an exposure to the rhythmic sequences measured as probability of occurrence or familiarity was examined and evaluated in a Bayesian framework. Using this framework, we have demonstrated the possibility of predicting rhythm perception from rhythm production data and priors. The method offers a new perspective on the relation between music rhythm perception and production;the rhythm perception process is adaptive to information in one's environment. Finally, by looking at the degree of transfer of learning on rhythm production, we have shown that certain rhythmic patterns seem to be represented in a more similar way than others. Therefore we have indirectly shown that the categorical representation of rhythmic patterns is involved in the process of production. The major findings from each chapter are reviewed below.

In Chapter 2, we studied the acculturation effect on rhythm production. We detected a cultural difference in the rhythm production of two-interval rhythmic patterns. Japanese percussion players tended to produce rhythmic patterns with less contrasted interval ratios than Dutch percussion players. However, this tendency holds only for the rhythmic ratios that consist of more contrasted durations, such as 1:4 and 1:5. The cultural difference was more salient for patterns that are assumed to yield more complex mental representations. The finding shows that the size of this acculturation effect interacts with complexity of the rhythms.

An empirical association between quantifiable properties of rhythmic contrasts in natural speech and written music was examined in Chapter 3. A measure of contrast of successive durations in rhythmic sequences called normalized Pairwise Variability Index (nPVI) was utilized. Patel and Daniele (2003) have shown that the rhythm of speech influences composed musical rhythm in instrumental and classical music in Western cultures (English and French). We extended this finding to a non-western language, Japanese, and to songs. Furthermore, a way to compute the nPVI from an underlying distribution assuming independence of two successive durations is proposed. Using this measure, it has been shown that nPVI differences between two cultures might reflect differences in the statistical distributions of durations chosen. Furthermore, the patterns of occurrence of short-long and long-short combination of durations in the rhythms were different between, but similar within, cultural groups. These observations held both for music and speech. The findings indicate that the rhythmic characteristics of natural speech have an influence on rhythms in composed music.

In Chapter 4 we demonstrated that a Bayesian approach provides an understanding of the difference that has been shown between perception and production of simple rhythmic patterns. The method formalizes the perceptual competition between mental representations of the rhythmic patterns and assumes possible non-uniform a priori probabilities of the rhythmic categories. Using this approach, the two kinds of information are related and we can predict perception data from production data. The contrast between rhythm perception and production data, taken from different studies, was shown to almost disappear, using independent prior probabilities from counts of patterns in corpora of musical scores, or from a theoretical measure of rhythmic complexity. The success of this Bayesian formalization may be interpreted as an optimal adaptation of our perceptual system to the environment in which the produced rhythms occur.

The issues examined in the previous chapters as well as the method proposed were combined in Chapter 5. Two issues are highlighted: the acculturation effect in relation to the complexity of the rhythms that was found in Chapter 2 and the possible association between speech and rhythm suggested in Chapter 3. To elaborate the concept of complexity of rhythmic patterning, two complexity measures of musical rhythm were identified and examined, namely, a syncopation measure, which represents hierarchical complexity, and the measure of the durational variability of successive intervals, which represents serial complexity. Perception, production and familiarity judgments of three interval rhythmic patterns were studied in combination. They were coupled in the Bayesian way that was proposed in Chapter 4, using familiarity judgments as prior knowledge of rhythmic categories.

Examination of the data of Dutch and Japanese pianists suggested that characteristics of all tasks are shared for simple temporal patterns, while diversity emerged for patterns that were complex in terms of their hierarchical structure. Furthermore, observed differences in rhythm production exhibited regularity in terms of the cultural background of the participants: The Dutch musicians produced shorter last intervals than the Japanese. This can be linked to a corresponding finding for speech, and we think that exposure to the rhythm of the mother tongue is indeed a plausible explanation of this difference. Furthermore, we demonstrated that, by incorporating familiarity, the different characteristics of the data for perception and production could be successfully translated and associated by means of a Bayesian relation.

Chapter 6 highlighted rhythm perception and production from a somewhat different perspective than the other chapters. The chapter was concerned with learning to perform short musical rhythms with expressive variations. The influence of rhythm production training (imitation) on production (imitation) and perception (discrimination) tasks was examined. The degree of transfer of learning varied with stimuli type, which provided insight into the issue of which aspects of expressive music performance share a common basis. Performance was improved for the imitation task but not for the discrimination task; transfer of learning takes place within domains (perception - production) but not across domains. This indicates that perception and production are not completely based on the same source, and indeed, as pointed out in the previous chapters, the nature of these two tasks may be very different. On the other hand, the skill to imitate expressive timing transfers more easily to rhythms that have a different surface structure but belong to the same categorical representation, than to rhythms that do not have a common categorical representation. This may indicate that different surface structures of rhythms characterized by a common categorical representation are closely connected and may even be processed in a similar way. Furthermore, transfer of learning does not behave in the same way for timing and for loudness, which suggests that these parameters are not likely to be integrated into a single dimension of musical expression.

This thesis aims at contributing to our knowledge of music rhythm perception and production. One of the major contributions of this thesis is in offering more evidence of an influence of exposure to environmental acoustical cues on one's rhythm perception and production. Although it remains difficult to generalize which aspect of speech rhythm actually influences which aspect of music rhythm perception and production, several cases of this cross-domain association reported in the chapters form a useful basis for the future research. Furthermore, we have gone beyond mere empirical observations. We have provided the computational framework of how our perception process incorporates environmental information.

At every waking moment, we are exposed to a variety of acoustical cues. It is plausible that acoustical rhythms in every day life influence how one perceives and produces rhythm. Such a 'sense of rhythm' may be partly shared between domains such as language and music. This is what in effect causes a Japanese person to perceive a 'rizumu' differently from a Dutch person listening to the same 'ritme'.

References

Adachi, M. (2004). Non-musician's expression of emotion in music: How do Japanese manipulate a familiar song? *Proceedings of the 8th International Conference on Music Perception and Cognition*, Evanston, 46-47.

Altenmüller, E., Schurmann, K., Lim, V. K., & Parlitz, D. (2002). Hits to the left, flops to the right: Different emotions during listening to music are reflected in cortical lateralisation patterns. *Neuropsychologia*, 40, 2242-2256.

Ashley, R. (2002). Do[n't] change a hair for me: The art of jazz rubato. *Music Perception*, 19, 311-332.

Askenfelt, A. (1986). Measurement of bow motion and bow force in violin playing. *Journal of the Acoustical Society of America*, 80, 1007-1015.

Barlow, H., & Morgenstern, S. (1948). *A dictionary of musical themes*. New York: Crown Publishers.

Barlow, H., & Morgenstern, S. (1983). *A dictionary of musical themes* (Rev. ed.). London: Faber & Faber.

Beckman, M., Edwards, J., & Fletcher, J. (1992). Prosodic structure and tempo in a sonority model of articulatory dynamics. In G. Docherty & D. R. Ladd (Eds.), *Laboratory phonology 2: Gesture, segment, prosody* (pp. 68-86). Cambridge: Cambridge University Press.

Bengtsson, I., & Gabrielsson, A. (1980). Methods for analyzing performance of musical rhythm. *Scandinavian Journal of Psychology*, 21, 257-268.

Berlyne, D. E. (1971). *Aesthetics and psychobiology*. New York: Appleton-Century-Crofts.

Besson, M., & Schön, D. (2003). Comparison between language and music. In I. Peretz & R. J. Zatorre (Eds.), *The cognitive neuroscience of music* (pp. 269-293). Oxford ; New York: Oxford University Press.

Besson, M., Faita, F., Peretz, I., Bonnel, A. M., & Requin, J. (1998). Singing in the brain: Independence of lyrics and tunes. *Psychological Science*, 9, 494-498.

Birkhoff, G. D. (1933). *Aesthetic measure.* Cambridge, MA: Harvard University Press.

Boselie, F., & Leeuwenberg, E. (1985). Birkhoff revisited: Beauty as a function of effect and means. *American Journal of Psychology*, 98, 1-39.

Carterette, E. C., & Kendall, R. A. (1999). Comparative music perception and cognition. In D. Deutsch (Ed.), *The psychology of music* (pp. 725-791). San Diego: Academic Press.

Cemgil, A. T., Desain, P., & Kappen, H. J. (2000). Rhythm quantization for transcription. *Computer Music Journal*, 24, 60-76.

Clarke, E. F. (1985). Structure and expression in rhythmic performance. In P. Howell, I. Cross, & R. West (Eds.), *Musical structure and cognition* (pp.209-236). London: Academic Press.

Clarke, E. F. (1985). Structure and expression in rhythmic performance. In P. Howell, I. Cross & R. West (Eds.), *Musical structure and cognition* (pp. 209-236). London: Academic press.

Clarke, E. F. (1987). Categorical rhythm perception: An ecological perspective. In A. Gabrielsson (Ed.), *Action and perception in rhythm and music*: *No. 55*, pp. 19-33): Royal Swedish Academy of Music.

Clarke, E. F. (1988). Generative principles in music performance. In J. A. Sloboda (Ed.), *Generative processes in music: The psychology of performance, improvisation, and composition* (pp. 1-26). Oxford: Claredon press.

Clarke, E. F. (1993). Imitating and evaluating real and transformed musical performances. *Music Perception*, 10, 317-341.

Clarke, E. F. (1999). Rhythm and timing in music. In D. Deutsch (Ed.), The psychology of music (2nd ed., pp. 473-500). San Diego: Academic Press.

Cohen, A., Ivry, R. I., & Keele, S. W. (1990). Attention and structure in sequence learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 16, 17-30.

Collier, G. L., & Collier, J. L. (2002). A study of timing in two Louis Armstrong solos. *Music Perception*, 19, 463-483.

Collier, G. L., & Wright, C. E. (1995). Temporal rescaling of simple and complex ratios in rhythmic tapping. *Journal of Experimental Psychology: Human Perception and Performance*, 21, 602-627.

Cutting, J. E., & Rosner, B. S. (1974). Categories and boundaries in speech and music. *Perception & Psychophysics*, 16, 564-570.

Cvitanovic, P., Shraiman, B., & Sönderberg, B. (1985). Scaling laws of mode locking sin circle maps. *Pysica Scripta*, 3, 263-270.

Dauer, R. M. (1983). Stress-timing and syllable-timing reanalyzed. *Journal of Phonetics*, 11, 51-62.

Dauer, R. M. (1987). Phonetic and phonological components of language rhythm. *Proceedings of the 11th International Congress of Phonetic Science,* 5, 447-450.

Dayan, P., Hinton, G. E., & Neal, R. M. (1995). The Helmholtz Machine. *Neural Computation*, 7, 889-904.

Demany, L., & Semal, C. (2002). Limits of rhythm perception. *Quarterly Journal of Experimental Psychology*, 55 (A), 643-657.

Desain, P., & Honing, H. (1992). *Music, mind and machine: Studies in computer music, music cognition and artificial intelligence.* Amsterdam: Thesis Publishers.

Desain, P., & Honing, H. (1994). Does expressive timing in music performance scale proportionally with tempo. *Psychological Research,* 56, 285-292.

Desain, P., & Honing, H. (1999). Computational Models of Beat Induction: The Rule-Based Approach. *Journal of New Music Research*, 28, 29-42.

Desain, P., & Honing, H. (2003). The formation of rhythmic categories and metric priming. *Perception*, 32, 341-365.

Desain, P., & Sadakata, M. (in preparation-a). Categorization in the perception of two-interval rhythmic patterns.

Desain, P., & Sadakata, M. (in preparation-b). Could the regression effect be an artifact? Perception-action differences explained by Bayes rule.

Desain, P., Honing, H., van Thienen, H., & Windsor, W. L. (1998). Computational Modeling of Music Cognition: Problem or Solution? *Music Perception*, 16, 151-166.

Dowling, W. J., & Harwood, D. L. (1986). *Music cognition*. Orlando: Academic Press.

Drake, C. (1993a). Influence of age and musical experience on timing and intensity variations in reproductions of short musical rhythms. *Psychologica Belgica*, 33, 217-228.

Drake, C. (1993b). Perceptual and performed accents in musical sequences. *Bulletin of the Psychonomic Society*, 31, 107-110.

Drake, C. (1993c). Reproduction of musical rhythms by children, adult musicians and adult non-musicians. *Perception & Psychophysics*, 53, 25-33.

Drake, C., & Bertrand, D. (2001). The quest for universals in temporal processing in music. *Annals of the New York Academy of Sciences*, 930, 17-27.

Drake, C., & El Heni, J. B. (2003). Synchronizing with music: Intercultural differences. *Neurosciences and Music*, 999, 429-437.

Drake, C., & Palmer, C. (1993). Accent structures in music performance. *Music Perception*, 10, 343-378.

Drake, C., & Palmer, C. (2000). Skill acquisition in music performance: Relations between planning and temporal control. *Cognition*, 74, 1-32.

Drake, C., Penel, A., & Bigand, E. (2000a). Tapping in time with mechanically and expressively performed music. *Music Perception*, 18, 1-23.

Drake, C., Penel, A., & Bigand, E. (Eds.). (2000b). Why musicians tap slower than nonmusicians. In P. Desain & L. Windsor (Eds). *Rhythm perception and production* (pp.245-249), Lisse: Swets & Zeitlinger B. V.

Eimas, P. D., Siqueland, E. R., Jusczyk, P., & Vigorito, J. (1971). Speech perception in infants. *Science*, 171, 303-306.

Eisler, H. (1976). Experiments on subjective duration 1868-1975: A collection of power function exponents. *Psychological Bulletin*, 83, 1154–1171.

Ericsson, K. A., Krampe, R. T., & Tesch-Römer, C. (1993). The role of deliberate practice in acquisition of expert performance. *Psychological Review*, 100, 363-406.

Essens, P. J. (1986). Hierarchical organization of temporal patterns. *Perception & Psychophysics*, 40, 69–73.

Essens, P. J., & Povel, D. J. (1985). Metrical and nonmetrical representations of temporal patterns. *Perception & Psychophysics*, 37, 1-7.

Fraisse, P. (1946). Contribution a l'étude du rhythme en tant que forme temporelle. *Journal de psychologie normale et pathologique*, 39, 283–304.

Fraisse, P. (1956). *Les structures rhythmiques*. Louvain: Publications Universitaires de Louvain.

Fraisse, P. (1982). Rhythm and tempo. In D. Deutsch (Ed.), *The psychology of music* (pp. 149–180). New York: Academic Press.

Fraisse, P. (1984). Perception and estimation of time. *Annual Review of Psychology*, 35, 1-36.

Frensch, P. A., & Rünger, D. (2003). Implicit learning. *Current Directions in Psychological Science*, 12, 13-18.

Friberg, A., & Sundström, A. (2002). Swing ratios and ensemble timing in jazz performance: Evidence for a common rhythmic pattern. *Music Perception*, 19, 333-349.

Friston, K. (2002). Functional integration and inference in the brain. *Progress in Neurobiology*, 68, 113-143.

Gabrielsson, A. (1974). Performance of rhythm patterns. *Scandinavian Journal of Psychology*, 15, 63-72.

Gabrielsson, A. (1987). Once again: The theme from Mozart's piano sonata in a major (K.331). In A. Gabrielsson (Ed.), *Action and perception in rhythm and music* (pp. 81-103). Stockholm: The Royal Swedish Academy of Music.

Gabrielsson, A. (1999). The performance of music. In D. Deutsch (Ed.), *The psychology of music* (pp. 501-602). San Diego: Academic Press.

Gabrielsson, A. (2003). Music performance research at the millennium. *Psychology of Music*, 31, 221-272.

Gabrielsson, A., Bengtsson, I., & Gabrielsson, B. (1983). Performance of musical rhythm in 3/4 and 6/8 meter. *Scandinavian Journal of Psychology*, 24, 193-213.

Geisler, W. S., & Kersten, D. (2002). Illusions, perception and Bayes. *Nature Neuroscience*, 5, 508-510.

González, D. L., & Piro, O. (1985). Symmetric kicked self-oscillators: Iterated maps, strange attractors, and symmetry of the phase locking Farey hierarchy. *Physical Review Letters*, 55, 17-20.

Grabe, E. (2004). Intonational variation in urban dialects of english spoken in the british isles. In P. Gilles & J. Peters (Eds.), *Regional variation in intonation* (pp. 9-31). Tuebingen, Niemeyer: Linguistische Arbeiten.

Grabe, E., & Low, E. L. (2002). Durational variability in speech and the rhythm class hypothesis. In C. Gussenhoven & N. Warner (Eds.), *Laboratory phonology 7* (pp. 515-546). Berlin; New York: Mouton de Gruyter.

Grabe, E., Post, B., & Watson, I. (1999). The acquisition of rhythmic patterns in English and French. *Proceedings of the 14th International Congress of Phonetic Sciences*, USA, 1201-1204.

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York: Wiley.

Hannon, E. E., & Trehub, S. E. (2005). Metrical categories in infancy and adulthood. *Psychological Science*, 16, 48-55.

Harwood, D. L. (1976). Universals in music - perspective from cognitive psychology. *Ethnomusicology*, 20, 521-533.

Hayes, W. L. & Winkler, R. L. (1970). *Statistics: probability, inference, and decision, volume II.* New York: Holt. Rinehart and Winston.

Heuer, H., & Schmidt, R. A. (1988). Transfer of learning among motor patterns with different relative timing. *Journal of Experimental Psychology: Human Perception and Performance*, 14, 241-252.

Helm van der P. A. (2000). Simplicity versus likelihood in visual perception: From surprisals to Precisals. *Psychological Bulletin*, 126, 770-800.

Hirata, Y. (2004). Computer assisted pronunciation training for native English speakers learning Japanese pitch and durational contrasts. *Computer Assisted Language Learning*, 17, 357-376.

Hirsh-Pasek, K., Nelson, D. G. K., Jusczyk, P. W., Cassidy, K. W., Druss, B., & Kennedy, L. (1987). Clauses are perceptual units for young infants. *Cognition*, 26, 269-286.

Hommel, B., Musseler, J., Aschersleben, G., & Prinz, W. (2001). The theory of event coding (tec): A framework for perception and action planning. *Behavioral and Brain Sciences*, 24, 849-878.

Honing, H. (1990). POCO: an environment for analyzing, modifying, and generating expression in music. Proceedings of the International Computer Music Conference (pp.364-368). San Francisco: Computer Music Association.

Honing, H. (2003). The final ritard: On music, motion, and kinematic models. *Computer Music Journal*, 27, 66-72.

Hoopen ten, G., Sasaki, T., Nakajima, Y., Remijn, G., Massier, B., Rhebergen, K. S., et al. (in press). Time-shrinking and categorical temporal ratio perception: Evidence for a 1:1 temporal category. *Music Perception*.

Huron, D. (1999). Highpoints: A study of melodic peaks by Zohar Eitan. *Music Perception*, 16, 257-264.
Huron, D. (2002). Music information processing using the Humdrum Toolkit: Concepts, examples, and lessons. *Computer Music Journal*, 26, 15-30.
Huron, D., & Ollen, J. (2003). Agogic contrast in french and english themes: Further support for patel and daniele (2003). *Music Perception*, 21, 267-271.
Ihre, A. (1992). Production and perception of rhythm patterns within one beat. Unpublished master's thesis, Leiden University, The Netherlands.
Iversen, J. R., Patel, A. D., & Ohgushi, K. (2004, April). Perception of nonlinguistic rhythmic stimuli by American and Japanese listeners. *Proceedings of the 18th International Congress on Acoustics*, Kyoto, 2721-2724.
Jastrow, J. (1890). *The time-relations of mental phenomena*. New York: N. D. C. Hodges.
Jensen, F. V. (2001). *Bayesian Networks and Decision Graphs.* New York: Springer.
Jusczyk, P. W., & Krumhansl, C. L. (1993). Pitch and rhythmic patterns affecting infants sensitivity to musical phrase structure. *Journal of Experimental Psychology: Human Perception and Performance*, 19, 627-640.
Kendall, R. A., & Carterette, E. C. (1990). The communication of musical expression. *Music Perception*, 8, 129-163.
Kilborn, K., & Ito, T. (1989). Sentence processing strategies in adult bilinguals. In B. MacWhinney & E. Bates (Eds.), *The crosslinguistic study of sentence processing* (pp. 257-291). Cambridge: Cambridge University Press.
Knill, D. C., & Richards, W. (Eds.). (1996). *Perception as Bayesian inference*. Cambridge: Cambridge University Press.
Krumhansl, C. L., & Jusczyk, P. W. (1990). Infants perception of phrase structure in music. *Psychological Science*, 1, 70-73.
Kubozono, H. (2002). Temporal neutralization in japanese. In C. Gussenhoven & N. Warner (Eds.), *Laboratory phonology 7* (pp. 171-201). Berlin; New York: Mouton de Gruyter.
Large, E. W., & Jones, M. R. (1999). The dynamics of attending: How people track time-varying events. *Psychological Review*, 106, 119-159.
Lempel, a., & Ziv, J. (1976). Complexity of finite sequences*. IEEE Transactions on Information Theory*, 22, 75-81.
Lerdahl, F., & Jackendoff, R. (1983). *A generative theory of tonal music*. Cambridge, Mass: MIT Press.
Ling, L. E., Grabe, E., & Nolan, F. (2000). Quantitative characterizations of speech rhythm: Syllable-timing in Singapore English. *Language and Speech*, 43, 377-401.
London, J. (2001). Rhythm. In *The New Grove Dictionary of Music and Musicians* (Rev. ed., Vol. 21, pp. 277-309). London: Macmillan.
Longuet-Higgins, H. C. (1987). *Mental Processes*. Cambridge, Mass: MIT Press.
Longuet-Higgins, H. C., & Lee, C. S. (1982). The perception of musical rhythms. *Perception*, 11, 115-128.
Longuet-Higgins, H. C., & Lee, C. S. (1984). The rhythmic interpretation of monophonic music. *Music Perception*, 1, 424-441.
MacDonald, P. A. R., & Wilson, G. B. (2006). Constructions of jazz: How jazz musicians present their collaborative musical practice. *Musicae Scientiae*, 10, 59-83.
Maess, B., Koelsch, S., Gunter, T. C., & Friederici, A. D. (2001). Musical syntax is processed in broca's area: An meg study. *Nature Neuroscience*, 4, 540-545.

Marcus, G. F., Vijayan, S., Rao, S. B., & Vishton, P. M. (1999). Rule learning by seven-month-old infants. *Science*, 283, 77-80.
Martin, J. G. (1972). Rhythmic (hierarchical) versus serial structure in speech and other behavior. *Psychological Review*, 79, 487-509.
Maye, J., Werker, J. F., & Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition*, 82, B101-B111.
McMullen, E., & Saffran, J. R. (2004). Music and language: A developmental comparison. *Music Perception*, 21, 289-311.
Meyer, R. K., & Palmer, C. (2003). Temporal and motor transfer in music performance. *Music Perception*, 21, 81-104.
MIDI Database. (n.d.). Retrieved May 31 2003, from *http://www.mididb.com/*
Mito, H., & Murao, T. (1999). Memory for Japanese pop songs with different styles: Role of combination of text with melody. In S. W. Yi (Ed.), *Music, mind and science* (pp. 393-407). Seoul: Seoul National University Press.
Nakamura, T. (1987). The communication of dynamics between musicians and listeners through musical performance. *Perception & Psychophysics*, 41, 525-533.
Nazzi, T., Bertoncini, J., & Mehler, J. (1998). Language discrimination by newborns: Toward an understanding of the role of rhythm. *Journal of Experimental Psychology: Human Perception and Performance*, 24, 756-766.
Nespor, M. (Ed.). (1990). *On the rhythm parameter in phonology*. Dordrecht: Foris.
New York: Clarendon Press.
Ohgushi, K. (2002). Comparison of dotted rhythm expression between Japanese and western pianists. *Proceedings of the 7th International Conference on Music Perception and Cognition*, Sydney, 250-253.
Osterhout, L., & Holcomb, P. J. (1993). Event-related potentials and syntactic anomaly - evidence of anomaly detection during the perception of continuous speech. *Language and Cognitive Processes*, 8, 413-437.
Palmer, C. (1989). Mapping musical thought to musical performance. *Journal of Experimental Psychology: Human Perception and Performance*, 15, 331-346.
Palmer, C. (1997). Music Performance. *Annual Review of Psychology*, 48, 115-138.
Palmer, C., & Drake, C. (1997). Monitoring and planning capacities in the acquisition of music performance skills. *Canadian journal of Exprimental Psychology*, 51, 369-384.
Palmer, C., & Kelly, M. H. (1992). Linguistic prosody and musical meter in song. *Journal of Memory and Language*, 31, 525-542.
Palmer, C., & Krumhansl, C. L. (1990). Mental representations of musical meter. *Journal of Experimental Psychology: Human Perception and performance*, 16, 728-741.
Palmer, C., & Meyer, R. K. (2000). Conceptual and motor learning in music performance. *Psychological Science*, 11, 63-68.
Parzen, E. (1962). Estimation of a probability density-function and mode. *Annals of Mathematical Statistics*, 33, 1065-1076.
Patel, A. D. (1998). Syntactic processing in language and music: Different cognitive operations, similar neural resources? *Music Perception*, 16, 27-42.
Patel, A. D. (2003). Language, music, syntax and the brain. *Nature Neuroscience*, 6, 674-681.
Patel, A. D., & Daniele, J. R. (2003a). An empirical comparison of rhythm in language and music. *Cognition*, 87, B35-B45.

Patel, A. D., & Daniele, J. R. (2003b). Stress-timed vs. Syllable-timed musk? - a comment on huron and ollen (2003). *Music Perception*, 21, 273-276.

Patel, A. D., Peretz, I., Tramo, M., & Labreque, R. (1998). Processing prosodic and musical patterns: A neuropsychological investigation. *Brain and Language,* 61, 123-144.

Penel, A., & Drake, C. (1998). Sources of timing variation in music performance: A psychological segmentation model. *Psychological Research*, 61, 12-32.

Penel, A., & Drake, C. (1998). Sources of timing variations in music performance: A psychological segmentation model. *Psychological Research*, 61, 12-32.

Penel, A., & Drake, C. (1999). Seeking "one" expressive timing. In S. W. Yi (Ed.), *Music, Mind and Science* (pp. 271-297). Seoul: Seoul University Press.

Peper, C. E., Beek, P. J., & van Wieringen, P. C. W. (1995). Multifrequency coordination in bimanual tapping: Asymmetrical coupling and signs of supercriticality. *Journal of Experimental Psychology: Human perception and performance*, 21, 1117-1138.

Peretz, I., & Coltheart, M. (2003). Modularity of music processing. *Nature Neuroscience*, 6, 688-691.

Peretz, I., Gagnon, L., Hebert, S., & Macoir, J. (2004a). Singing in the brain: Insights from cognitive neuropsychology. *Music Perception*, 21, 373-390.

Peretz, I., Radeau, M., & Arguin, M. (2004b). Two-way interactions between music and language: Evidence from priming recognition of tune and lyrics in familiar songs. *Memory & Cognition*, 32, 142-152.

Pitts, S. E. (2002). Changing tunes: Musical experience and self-perception amongst school and university music students. *Musicae Scientiae*, 6, 73-92.

Povel, D. J. (1981). Internal representation of simple temporal patterns. *Journal of Experimental Psychology: Human Perception and Performance*, 7, 3–18.

Povel, D. J. (1984). A theoretical framework for rhythm perception. *Psychological Research*, 45, 315-337.

Povel, D. J., & Essens, P. (1985). Perception of temporal patterns. *Music Perception*, 2, 411-440.

Povel, D. J., & Okkerman, H. (1981). Accents in equitone sequences. *Perception & Psychophysics*, 30, 565-572.

Pressing, J. (n.d.). Cognitive complexity and the structure of musical patterns. Retrieved August 29, 2003, from *http://psy.uq.edu.au/CogPsych/Noetica/OpenForumIssue8/Pressing.html*

Ramus, F. (2002). Acoustic correlates of linguistic rhythm: Perspectives. *Proceedings of the Speech Prosody, Université de Provence*. France, 115-120.

Ramus, F., Nespor, M., & Mehler, J. (1999). Correlates of linguistic rhythm in the speech signal. *Cognition*, 73, 265-292.

Repp, B. H. (1984). Categorical perception: Issues, methods, findings. In N. J. Lass (Ed). *Speech and Language: Advances in basic research and practice, Vol. 10* (pp.243-335). New York: Academic Press.

Repp, B. H. (1990). Patterns of expressive timing in performances of a beethoven minuet by 19 famous pianists. *Journal of the Acoustical Society of America*, 88, 622-641.

Repp, B. H. (1992a). Diversity and commonality in music performance - an analysis of timing microstructure in Schumann's Traumerei. *Journal of the Acoustical Society of America*, 92, 2546-2568.

Repp, B. H. (1992b). Probing the cognitive representation of musical time - structural constraints on the perception of timing perturbations. *Cognition*, 44, 241-281.
Repp, B. H. (1994). Relational invariance of expressive microstructure across global tempo changes in music performance - an exploratory-study. *Psychological Research*, 56, 269-284.
Repp, B. H. (1995a). Detectability of duration and intensity increments in melody tones - a partial connection between music perception and performance. *Perception & Psychophysics*, 57, 1217-1232.
Repp, B. H. (1995b). Expressive timing in schumann traumerei - an analysis of performances by graduate student pianists. *Journal of the Acoustical Society of America*, 98, 2413-2427.
Repp, B. H. (1997). Expressive timing in a debussy prelude: A comparison of student and expert pianists. *Musicae Scientiae*, 1, 257-268.
Repp, B. H. (1998). A microcosm of musical expression. I. Quantitative analysis of pianists' timing in the initial measures of Chopin's etude in e major. *Journal of the Acoustical Society of America*, 104, 1085-1100.
Repp, B. H. (1998). Variation on a theme by Chopin: Relations between perception and production of timing in music. *Journal of Experimental Psychology: Human Perception and Performance*, 24, 791-811.
Repp, B. H. (1999a). A microcosm of musical expression: II. Quantitative analysis of pianists' dynamics in the initial measures of Chopin's etude in e major. *Journal of the Acoustical Society of America*, 105, 1972-1988.
Repp, B. H. (1999b). A microcosm of musical expression. III. Contributions of timing and dynamics to the aesthetic impression of pianists' performances of the initial measures of Chopin's etude in e major. *Journal of the Acoustical Society of America*, 106, 469-478.
Repp, B. H. (1999c). Detecting deviations from metronomic timing in music: Effects of perceptual structure on the mental timekeeper. *Perception & Psychophysics*, 61, 529-548.
Repp, B. H. (1999d). Relationships between performance timing, perception of timing perturbations, and perceptual-motor synchronisation in two Chopin preludes. *Australian Journal of Psychology*, 51, 188-203.
Repp, B. H., Windsor, W. L., & Desain, P. (2002). Effects of tempo on the timing of simple musical rhythms. *Music Perception*, 19, 563-591.
Rizzolatti, G., Fadiga, L., Gallese, V., & Fogassi, L. (1996). Premotor cortex and the recognition of motor actions. *Cognitive Brain Research*, 3, 131-141.
Roach, P. (1982). On the distinction between 'stress-timed' and 'syllable-timed' languages. In D. Crystal (Ed.), *Linguistic controversies* (pp. 73-79). London: Edward Arnold.
Roberts, S., & Pashler, H. (2000). How persuasive is a good fit? A comment on theory testing. *Psychological Review*, 107, 358-367.
Robertson, I. (2000). Imitative problem solving: Why transfer of learning often fails to occur. *Instructional Science*, 28, 263-289.
Sadakata, M., & Desain, P. (submitted-a). Comparing rhythmic structure in popular music and speech.
Sadakata, M., & Desain, P. (submitted-b). Diversity and commonality in rhythm perception and production: An influence of rhythmic complexity and culture.

Sadakata, M., Desain, P., & Honing, H. (2006). The Bayesian way to relate rhythm perception and production. *Music Perception*, 23, 267-286.

Sadakata, M., Ohgushi, K., & Desain, P. (2004). A cross-cultural comparison study of the production of simple rhythmic patterns. *Psychology of Music*, 32, 389-404.

Saito, H. (1999). *Saito Hideo kougi-roku [the transcript of lecutres given by hideo saito]*. Tokyo: Hakusuisha.

Samuel, a. G. (1981). Phonemic restoration - insights from a new methodology. *Journal of Experimental Psychology: General*, 110, 474-494.

Schaffrath, H. (1993). Repräsentation einstimmiger Melodien: computerunterstützte Analyse und Musikdatenbanken. In B. Enders & S. Hanheide (Eds.), *Neue Musiktechnologie* (pp. 277-300), Mainz: Schott.

Schaffrath, H. (1995). The Essen Folksong Collection in the Humdrum Kern Format [Computer database]. D. Huron (Ed.). Menlo Park, CA: Center for Computer Assisted Research in the Humanities.

Schulze, H. H. (1989). Categorical perception of rhythmic patterns. *Psychological Research*, 51, 10-15.

Scruggs, T. E., & Mastropieri, M. A. (1988). Acquisition and transfer of learning-strategies by gifted and nongifted students. *Journal of Special Education*, 22, 153-166.

Seashore, C. E. (1938). *Psychology of music* (1st ed.). New York, London,: McGraw-Hill Book Company inc.

Senju, M., & Ohgushi, K. (1987). How are the players ideas conveyed to the audience. *Music Perception*, 4, 311-323.

Serafine, M. L., Crowder, R. G., & Repp, B. H. (1984). Integration of melody and text in memory for songs. *Cognition*, 16, 285-303.

Serafine, M. L., Davidson, J., Crowder, R. G., & Repp, B. H. (1986). On the nature of melody text integration in memory for songs. *Journal of Memory and Language,* 25, 123-135.

Shaffer, L. H. (1984). Timing in solo and duet piano performances. *Quarterly Journal of Experimental Psychology Section*, 36 (A), 577-595.

Shaw, M., & Coleman, H. (1960). *National Anthems of the World*. London: Pitman.

Shibata, M. (1983). *Nihon no oto wo kiku [listening to the Japanese sounds]*. Tokyo: Seidosha.

Shmulevich, I., & Povel, D. J. (2000). Complexity measures of musical rhythms. In P. Desain & L. Windsor (Eds.), *Rhythm perception and production* (pp. 239-244). Lisse, NL: Swets & Zeitlinger.

Siegel, J. A., & Siegel, W. (1977). Categorical perception of tonal intervals - musicians cant tell sharp from flat. *Perception & Psychophysics*, 21, 399-407.

Sloboda, J. A. (1982). Music performance. In D. Deutsch (Ed.), *The psychology of music* (pp. 479-496). New York: Academic Press.

Sloboda, J. A. (1985). Expressive skill in two pianists - metrical communication in real and simulated performances. *Canadian Journal of Psychology*, 39, 273-293.

Sloboda, J. A. (1985). *The musical mind: The cognitive psychology of music*. Oxford: Oxford University Press.

Sloboda, J. A., Davidson, J. W., Howe, M. J. A., & Moore, D. G. (1996). The role of practice in the development of performing musicians. *British Journal of Psychology*, 87, 287-309.

Sternberg, S., Knoll, R. L., & Zukofsky, P. (1982). Timing by skilled musicians. In D. Deutsch (Ed.), *The psychology of music* (pp. 181–239). New York: Academic Press.

Stevens, C. (2004, April). The effect of first language tonality on perception of pitch in speech and music. *Proceedings of the 18th International Congress on Acoustics*, Kyoto, 2717-2720.

Stobart, H., & Cross, I. (2000). The Andean anacrusis? Rhythmic structure and perception in Easter songs of northern Potosi, Bolivia. *British Journal of Ethnomusicology*, 9, 63-94.

Studdert, M., Liberman, a. M., Harris, K. S., & Cooper, F. S. (1970). Theoretical notes motor theory of speech perception. *Psychological Review*, 77, 234-249.

Summers, J. J., Bell, R., & Burns, B. D. (1989). Perceptual and motor factors in the imitation of simple temporal patterns. *Psychological Research*, 51, 23–27.

Summers, J. J., Hawkins, S. R., & Mayers, H. (1986). Imitation and production of interval ratios. *Perception & Psychophysics*, 39, 437–444.

Sundberg, J., & Verrillo, V. (1980). On the anatomy of the retard - a study of timing in music. *Journal of the Acoustical Society of America*, 68, 772-779.

Sundberg, J., Prame, E., & Iwarsson, J. (1996). Replicability and accuracy of pitch patterns in professional singers. In P. Davis & N. Fletcher (Eds.), *Vocal fold psychology, controlling complexity and chaos* (pp. 291-306). San Diego: CA: Singular Publishing Group.

Tanguiane, A. S. (1993). *Artificial perception and music recognition*. Berlin: Springer-Verlag.

Tanner, W. P., & Swets, J. A. (1954). A decision-making theory of visual detection. *Psychological Review*, 61, 401-409.

Temperley, D. (2002). A Bayesian Approach to key-finding. In C. Anagnostopoulou, M. Ferrand, & A. Smaill (Eds.), Music and Artificial Intelligence (pp. 195-206). Berlin: Springer-Verlag.

Tervaniemi, M., Medvedev, S. V., Alho, K., Pakhomov, S. V., Roudas, M. S., van Zuijen, T. L., et al. (2000). Lateralized automatic auditory processing of phonetic versus musical information: A pet study. *Human Brain Mapping*, 10, 74-79.

Timmers, R. (2002). *Freedom and constraints in timing and ornamentation*. Maastricht: Shaker Publishng BV.

Timmers, R. (2005). Predicting the similarity between expressive performances of music from measurements of tempo and dynamics. *The Journal of Acoustical Society of America*, 117, 391-399.

Todd, N. (1985). A model of expressive timing in tonal music. *Music Perception*, 3, 33-57.

Travlos, A. K. (1999). More practice does not necessarily enhance transfer of learning: Evidence and interpretations. *Perceptual and Motor Skills*, 89, 1161-1175.

Weiss, Y., Simoncelli, E. P., & Adelson, E. H. (2002). Motion illusions as optimal percepts. *Nature Neuroscience*, 5, 598-604.

Wenk, B. J. (1987). Just in time: On speech rhythms in music. *Linguistics*, 25, 969-981.

Windsor, W. L., & Clarke, E. F. (1997). Expressive timing and dynamics in real and artificial musical performance: Using an algorithm as an analytical tool. *Music Perception*, 15, 127-152.

Windsor, W. L., Desain, P., Penel, A., & Borkent, M. (2006). A structurally guided method for the decomposition of expression in music performance. *Journal of the Acoustical Society of America*, 119, 1182-1193.

Witt de, L. A., & Samuel, a. G. (1990). The role of knowledge-based expectations in music perception - evidence from musical restoration. *Journal of Experimental Psychology: General*, 119, 123-144.

Woody, R. H. (2000). Learning expressivity in music performance: An exploratory study. *Research Studies in Music Education*, 14, 14-23.

Samenvatting / Summary in Dutch

Dit proefschrift draagt bij aan onze kennis over het waarnemen en produceren van muzikale ritmes. De voornaamste bijdrages zijn nieuwe observaties van de invloed van onze klinkende omgeving op ritme perceptie en productie. Hiernaast wordt een theoretisch kader geschetst hoe perceptuele processen door deze informatie uit de omgeving worden beïnvloed. Dit kader kan voor toekomstig onderzoek een basis bieden bij het testen van hypotheses over de relatie tussen perceptie en productie van ritmes. De belangrijkste bevindingen van elk hoofdstuk worden hieronder genoemd.

In Hoofdstuk 2 wordt onderzoek beschreven naar de invloed van culturele achtergrond op ritmeproductie. We observeerden dat Japanse percussionisten ritmes van twee intervallen met een minder gecontrasteerde intervalverhouding spelen dan Nederlandse percussionisten, wanneer de verhouding van de interval-lengtes toeneemt. Dit duidt aan dat de grootte van dit cultuur-effect afhankelijk is van de complexiteit van de ritmes, en groter wordt bij toenemende complexiteit.

Een empirisch waargenomen verband tussen kwantificeerbare eigenschappen van ritmische contrasten van genoteerde muziek en gesproken taal wordt in Hoofdstuk 3 besproken. In eerder werk werd aangetoond dat het ritme van spraak samenhangt met het muzikale ritme van instrumentale klassieke muziek in Westerse culturen (Engelse en Franse taal), door de variabiliteit van de lengtes van opeenvolgende noten te beschouwen. Deze bevinding hebben wij uitgebreid naar een niet-Westerse taal, Japans, en naar gezongen muziek. Ook vonden wij dat de waargenomen verschillen in deze variabiliteit tussen Japans en Engels, eigelijk voornamelijk verschillen zijn in de onderliggende statistische verdeling van tijdsduren in zowel muziek als spraak. Dit ondersteunt de bevindingen van een samenhang van spraakritme met gecomponeerd muzikaal ritme.

In Hoofdstuk 4 wordt getoond hoe met een Bayesiaanse benadering de relatie beschreven kan worden tussen perceptie en productie van eenvoudige ritmische patronen. De methode formaliseert de perceptuele competitie tussen alternatieve mentale representaties van de ritmische patronen en baseert zich op mogelijke non-uniforme a priori waarschijnlijkheden van ritmische categorieën. Met deze methode worden twee soorten informatie aan elkaar gerelateerd en kunnen karakteristieken van ritmeperceptie worden voorspeld uit ritmeproductie. Het succes van deze Bayesiaanse formalisatie kan geïnterpreteerd worden als een optimale aanpassing van ons perceptuele systeem aan de omgeving waarin deze ritmes voorkomen.

De onderwerpen uit voorgaande hoofdstukken worden in Hoofdstuk 5 gecombineerd. Drie van deze onderwerpen worden eruit gelicht: het effect van culturele achtergrond bij toenemende complexiteit van ritmes uit Hoofdstuk 1, het mogelijke verband tussen de ritmes van gesproken taal en muziek uit Hoofdstuk 2 en de Bayesiaanse methode om ritmeperceptie te relateren aan ritmeproductie uit Hoofdstuk 4. Onderzoeksdata van Nederlandse en Japanse pianisten suggereert dat voor simpele temporele patronen, culturele achtergrond geen effect heeft, terwijl bij structureel complexere patronen er verschillen zichtbaar worden. De waargenomen verschillen in ritmeproductie toonden een regelmatig patroon afhankelijk van de culturele achtergrond van de proefpersoon. Deze bevinding zijn te interpreteren als analoog aan elders gerapporteerde bevindingen voor gesproken taal. Hiernaast hebben we aangetoond dat, met bekendheid met de ritmes als onderdeel van de ondezoeksopzet, de verschillende eigenschappen van perceptie en productie goed beschreven kunnen worden met een Bayesiaans verband.

In Hoofdstuk 6 wordt de invloed onderzocht van training van ritmeproductie op zowel ritmeproductie als ritmeperceptie. Uit eerder werk blijkt dat ritmes met verschillende temporele patronen of accenten in luidheid, worden waargenomen als hetzelfde patroon. De vaardigheid om ritmes met expressieve accenten te imiteren lijkt makkelijker te generaliseren naar ritmische patronen die in dezelfde categorische representatie vallen dan naar andere perceptuele categorieën. Dit zou kunnen wijzen op een nauwe verbinding en mogelijke gemeenschappelijke verwerking van verschillende ritmes die in dezelfde categorie vallen. Generalizatie van het training effect bleek niet hetzelfde voor temporele verschillen als voor verschillen in luidheid, wat suggereert dat deze muzikale dimensies niet geïntegreerd worden verwerkt. Hiernaast bleek ook dat overdracht van training gemakkelijker verloopt binnen één domein (i.e. perceptie) dan tussen domeinen (i.e. van productie naar perceptie). Dit wijst erop dat perceptie en productie niet op dezelfde mechanismen gebaseerd zijn, en zoals in de eerdere hoofdstukken besproken is, zou er wellicht een substantiëel verschil kunnen zijn in de aard van deze twee vaardigheden.

De woorden *Ritme* [rItmə] en *Rizumu* [ɹizumu] verwijzen naar ritme in het Nederlands en Japans. Deze woorden werden voor de titel gekozen om nadruk te leggen op de prosodische eigenschappen van verschillende talen. Op ieder moment worden wij blootgesteld aan verschillende klinkende fenomenen en het is plausibel te denken dat deze akoestische omgeving van alledag van invloed is op hoe wij ritme waarnemen en produceren. Dit 'ritmegevoel' zou gemeenschappelijk kunnen zijn voor verschillende domeinen, zoals taal en muziek. Het gevolg hiervan is dat bijvoorbeeld iemand met Japans als moedertaal een 'rizumu' iets anders hoort dan een Nederlander die hetzelfde 'ritme' hoort.

Translated by Rebecca Schaefer

研究のまとめ / Summary in Japanese

本研究では、音楽におけるリズムの知覚と産出に関して考察した。第一章はイントロダクション、第二章から第六章はそれぞれが独立した研究報告である。本研究の貢献は、次の二点に要約することができる。1）日常の音環境が個人のリズムの知覚と産出に影響を及ぼしていることを示唆する新たな事例を報告したこと、2）音環境情報がリズムの知覚過程にどのように組み込まれているかに関する理論的枠組みを提案したこと、である。2）の理論的枠組みはおそらく、今後、リズムの知覚産出過程に関する新しい仮説を実証するための基礎として役立つであろう。以下、本研究のまとめを章別に記す。

第二章では、リズム産出における文化の影響について考察した。対照的な比率をもつ二音リズムの産出課題において、日本人打楽器奏者はオランダ人打楽器奏者よりも小さな（二音の長さの比率が均一に近づくような傾向の）比率で演奏することが実験により示された。リズム産出における文化差の効果は、複雑な心的表象を必要とすると思われる比較的複雑な構造を有するリズムにおいてのみ認められた。この発見は、文化差の程度がリズム構造の複雑性と相関していることを示している。

第三章では、スピーチリズムと音楽リズムの関連について定量的に考察した。Patel & Daniele (2003)は、隣り合わせる音の長さの変化度合いに注目し、西洋器楽音楽の楽譜リズムが作曲家の母国語のスピーチリズムと相関していることを示した。我々は彼らの発見を非西洋の言語と歌曲（日本語）にまで拡大した。さらに、リズムを構成する音の長さの統計上の分布が文化グループ（英語と日本語のグループ）により異なることが、この章で認められた文化差を反映していることが示唆された。これらの発見は、スピーチリズムが音楽リズムに影響を及ぼしている、という主張をさらに強化するものである。

第四章では、二音リズムパタンの産出と知覚について、ベイズの定理を用いて関係づけることを提案した。このアプローチにおける重要な前提は次の二点である：1）リズム知覚過程において、一つのリズムが提示されたときに、複数のリズムカテゴリが「知覚されるカテゴリ」の候補として心的に競合している、2）競合しているリズムカテゴリのプライヤ確率は同等ではない。 我々はこれらの前提をもとに、ベイズの定理を用いてリズムの産出データとプライヤ確率とを組み合わせると、知覚データパタンをかなりよく予測することができることを示した。このアプローチは、我々の知覚システムが様々な事象の起こる日常環境に順応していることを定式化したものである。実測データとモデルデータとの高い相関は、このモデルがリズムの知覚産出過程をうまく説明できていることを示唆しているかもしれない。

第五章では、第二、三、四章で述べてきた、三つの事柄について総合的に考察した。それぞれ、第二章で述べた文化差とリズムの複雑性の相関の問題、第三章で考察した言語と音楽のリズムの関連について、第四章で提案したリズム知覚と産出をベイズの定理によって関連づける試みについて、である。 オランダ人と日本人のピアノ奏者のリズム知覚、産出、熟知度について調べた結果、複雑なリズムパタンの産出傾向が文化グループによって系統的に異なることが 示された。この系統的な文化差は、スピーチの知覚について示された文化差と関連づけられそうである。さらに、ベイズの定理によってあるリズムの産出データとプライヤ確率（この章ではリズムの熟知度をプライヤ確率として用いた）とを組み合わせることで、知覚データをかなり精密に再現することができることを示した。

第六章では、リズムの産出と知覚における、産出トレーニングの影響、つまり学習の移転効果について考察した。主要な発見三点について要約する。１）厳密には異なるタイミングやラウドネスを有するリズムが、同じカテゴリに属すると知覚されることはよく知られている。リズムを産出する技術の学習効果は、異なるカテゴリに属するリズムよりも、同じカテゴリに属するリズムにより簡単に移転するようである。このことは、同じカテゴリに属するリズムパタンが、心的に近い関係として認識されていること、そしてさらには似たような過程を経て情報処理されているかもしれないことを示唆する。２）タイミングに関する学習移転とラウドネスに関する学習移転とを分けて調べてみたところ、学習の転移パタンはタイミングとラウドネスとで異なることが示された。このことは、この二つの要素が、音楽表現の単一次元の構成要素ではないことを示している。３）学習された技術は、同一タスク内（産出から産出へ）のほうが、異なるタスク間（産出から知覚へ）よりも比較的容易に転移することが示された。このことは、産出と知覚タスクとが、まったく同一の基盤を有している訳ではないことを示唆する。他の章で指摘してきたように、両タスクの性質は異なるかもしれない。

本研究のタイトル、RitmeとRizumuとはそれぞれオランダ語と日本語でリズム、という意味であり、異なる言語間における独特の音韻的特徴を象徴している。どんな瞬間にも、我々は様々な音響情報にさらされている。日常の生活に存在する音響リズムが、個人のリズム産出知覚プロセスに影響を及ぼしているであろうことは想像に難くない。そのようにして培われた「リズム感」は音楽と言語のような、異なった分野において部分的に共用されているかもしれない。例えば、日本人の知覚する「リズム」とオランダ人の知覚する「ritme」とは違うかもしれないのである。

Dankwoord/Acknowledgment

Days of new experiences in a new land … the past four years have been so quick. I would like to thank the people who supported me.

First of all, I would like to thank Peter Desain for his daily supervision, support, enthusiasm, ideas, encouragement, kindness, and patience. I would also like to thank Halord Bekkering, Herbert Schriefers, and Charles de Weert for their support in providing a position and financial ground for me. I am grateful to the NWO and STW for their financial support. I am also grateful to the Rohm Music Foundation for awarding me a scholarship. It was very special to have the opportunity to conduct experiments at the Kyoto City University of Arts with help of Kengo Ohgushi and at the Utrecht Conservatorium with help of Henk Ekkel and Hans Timmermans. I appreciate the advice, ideas and help on chapters from Bret Aarden, Nozomi Azechi, Ali Taylan Cemgil, Anne Cutler, Ton Dijkstra, Esther Grabe, Bill Hartmann, Henkjan Honing, David Huron, John Iverson, Bert Kappen, Eric Kellerman, John A. Michon, Kengo Ohgushi, Ani D. Patel, Franck Ramus, Bruno Repp, Rebecca Shaefer, Renee Timmers, Dirk Vorberg, and Luke Windsor. Alex Brandmeyer and Paul Trilsbeek helped me with the setup of the experiments. Many people helped me with English edit on chapters, Alex Brandmeyer, Amanda Brown, Eric Kellerman, Martin Price, Simone Reis and Rebecca Scheafer. I also would like to thank Nozomi Azechi, Toshie Matsui and Kim Verhoef for their suggestions on non-English texts. Especially I would like to thank Rebecca Schaefer for translating the samenvatting. Keiko Sato gave me a valuable comment on the cover design of the thesis.

New and old members of Music Mind Machine group as well as people from NICI have provided me with great support and are a foundation of my research, Alex, Anja, Bass, David, Henkjan, Inge, Jenia, Kristin, Paul, Philip, Rebecca, Renee, Stéphane, Svetlana, Ton, Torsten, and Yvonne. Finally, I would like to thank for the support from family Kisters as well as from my family in Kyoto.

Publication list

Journal Papers

Sadakata, M., Ohgushi, K. & Desain, P. (2004). A cross-cultural comparison study of the production of simple rhythmic patterns. *Psychology of Music.* Vol. 32, No. 4, 389-403.

Sadakata, M., Desain, P., & Honing, H. (2006). Bayesian way to relate rhythm perception and production. *Music Perception. 23,* 267-286.

Hoppe, D., Sadakata, M., & Desain, P. (2006). Development of real-time visual feedback assistance in singing training: a review*. Journal of computer assisted learning.* Vol. 22, 308-316.

Sadakata, M., Desain, P. Correlating rhythm in speech and popular music: A cross-cultural comparison. Manuscript submitted for publication.

Sadakata. M., & Desain, P. Diversity and commonality in music rhythm perception and production: an influence of rhythmic complexity and culture. Manuscript submitted for publication.

Sadakata, M., Hoppe, D., & Desain, P. Learning to perform short musical rhythms with expressive deviations. Manuscript submitted for publication.

Conference papers

Sadakata, M., Ohgushi, K. & Desain, P. (2002). A cross-cultural comparison study of the production of simple rhythmic patterns. *Proceedings of International Conference of Auditory Display 2002 RenCon Workshop*.

Sadakata, M., Desain, P. & Honing, H (2002). The relation between rhythm perception and production: towards a Bayesianmodel. *Transaction of Technical Committee of Psychological and Physiological Acoustics*, Acoustical Society of Japan, 32 (10), H-2002-92.

Sadakata, M., Desain, P. & Honing, H, Patel, D. A., & Iverson, J. (2003). An analysis of rhythm in Japanese and English popular music. *Proceedings of The Japanese Society for Music Perception and Cognition.*

Sadakata, M., Desian, P., & Honing, H. (2004). An analysis of rhythmic ratios in scores of various kinds of music. *Proceedings of International Conference for Music Perception and Cognition.*

Sadakata, M. Desian, P., Honing, H., Patel, A. D. & Iversen, J. R. (2004). A cross-cultural study of the rhythm in English and Japanese popular music. *Proceedings of the International Symposium on Musical Acoustics*, Nara, 41-44.

Sadakata, M., Brandmeyer, A., Hoppe, D., Timmers, R., and Desain, P. (2006). Learning expressive performance of short musical rhythms with real-time visual feedback. *Proceedings of the Second International Conference on Music and Gesture*, 24.

Sadakata, M., Hoppe, D., & Desain, P. (2006). Learning to perform musical rhythms with expressive timing. *Proceedings of the Teaching, Learning and Performing music conference*.

Sadakata, M. & Desian, P. (2006). Perception and production of short western musical rhythms. *Proceedings of the International Conference for Music Perception and Cognition*.

Brandmeyer, A., Hoope, D., Sadakata, M., Timmers, R., & Deasin, P. (2006). PracticeSpace: A platform for real-time visual feedback in music instruction. *Proceedings of the International Conference for Music Perception and Cognition*.

Hoppe, D., Brandmeyer, A., Sadakata, M., Timmers, R., Desain, P. (2006). The effect of real-time visual feedback on the training of expressive performance skills. *Proceedings of the International Conference for Music Perception and Cognition*.

Curriculum Vitae

Makiko Sadakata was born on 29 May 1976 in Osaka, Japan. After graduating from Doshisha high school in Kyoto, she studied music composition at Kyoto City University of Arts (KCUA) from 1995 to 1999. During the study at KCUA, she met Kengo Ohgushi, who introduced her to the field of music cognition study. She followed the master program on music psychology at KCUA from 1999 to 2002 and received the master degree of musicology/music psychology at KCUA in 2002. While studying in Kyoto, she served as a secretary of the Japanese Society of Music Perception and Cognition. In 2000, she met Peter Desain, who was visiting KCUA. She visited him at the Music Mind Machine group, Nimegen Institute for Cognition and Information (NICI) in 2000 and 2001 to conduct experiments on the cross-cultural study of music rhythm production. After her master study, she applied for the junior researcher position at the same group in 2002, which became the first year of her doctoral research project. She was awarded a scholarship from the Rohm Music Foundation for the years 2004 and 2005.